National Library of Canada
Bibliothèque nationale du Canada

Canadian Theses Service
Service des thèses canadiennes

Ottawa, Canada
K1A 0N4

NOTICE

The quality of this microform is heavily dependent upon the quality of the original thesis submitted for microfilming. Every effort has been made to ensure the highest quality of reproduction possible.

If pages are missing, contact the university which granted the degree.

Some pages may have indistinct print especially if the original pages were typed with a poor typewriter ribbon or if the university sent us an inferior photocopy.

Previously copyrighted materials (journal articles, published tests, etc.) are not filmed.

Reproduction in full or in part of this microform is governed by the Canadian Copyright Act, R.S.C. 1970, c. C-30.

AVIS

La qualité de cette microforme dépend grandement de la qualité de la thèse soumise au microfilmage. Nous avons tout fait pour assurer une qualité supérieure de reproduction.

S'il manque des pages, veuillez communiquer avec l'université qui a conféré le grade.

La qualité d'impression de certaines pages peut laisser à désirer, surtout si les pages originales ont été dactylographiées à l'aide d'un ruban usé ou si l'université nous a fait parvenir une photocopie de qualité inférieure.

Les documents qui font déjà l'objet d'un droit d'auteur (articles de revue, tests publiés, etc.) ne sont pas microfilmés.

La reproduction, même partielle, de cette microforme est soumise à la Loi canadienne sur le droit d'auteur, SRC 1970, c. C-30.

NL-339 (r. 88/04)

Canada

THE UNIVERSITY OF ALBERTA

Expert Systems Applications in Hydrogeology

by

© Gordon Lee McClymont

A THESIS

SUBMITTED TO THE FACULTY OF GRADUATE STUDIES AND RESEARCH

IN PARTIAL FULFILMENT OF THE REQUIREMENTS FOR THE DEGREE

OF Master of Science

DEPARTMENT OF GEOLOGY

EDMONTON, ALBERTA

SPRING, 1988

Permission has been granted to the National Library of Canada to microfilm this thesis and to lend or sell copies of the film.

The author (copyright owner) has reserved other publication rights, and neither the thesis nor extensive extracts from it may be printed or otherwise reproduced without his/her written permission.

L'autorisation a été accordée à la Bibliothèque nationale du Canada de microfilmer cette thèse et de prêter ou de vendre des exemplaires du film.

L'auteur (titulaire du droit d'auteur) se réserve les autres droits de publication; ni la thèse ni de longs extraits de celle-ci ne doivent être imprimés ou autrement reproduits sans son autorisation écrite.

ISBN 0-315-42737-X

THE UNIVERSITY OF ALBERTA

RELEASE FORM

NAME OF AUTHOR: Gordon Lee McClymont

TITLE OF THESIS: Expert Systems Applications in Hydrogeology

DEGREE: Master of Science

YEAR THIS DEGREE GRANTED: Spring, 1988

Permission is hereby granted to THE UNIVERSITY OF ALBERTA LIBRARY to reproduce single copies of this thesis and to lend or sell such copies for private, scholarly or scientific research purposes only.

The author reserves other publication rights, and neither the thesis nor extensive extracts from it may be printed or otherwise reproduced without the author's written permission.

Gordon M. Clymont

#329, 27 Glenway Place SW
Calgary, Alberta
T3E 4T6

Date: December 10 1987

THE UNIVERSITY OF ALBERTA

FACULTY OF GRADUATE STUDIES AND RESEARCH

The undersigned certify that they have read, and recommend to the Faculty of Graduate Studies and Research for acceptance, a thesis entitled Expert Systems Applications in Hydrogeology submitted by Gordon Lee McClymont in partial fulfilment of the requirements for the degree of Master of Science.

Frank W. Schwartz
.....
Supervisor

.....

.....

.....

Date: December 10 1987

ABSTRACT

Expert systems are computer programs that assist users in solving problems by applying human expertise encoded into the program. A computer system called Expert ROKEY has been constructed that includes two expert systems (EXPAR and EXINS) and a contaminant transport model (ROKEY). ROKEY is an analytical code that accounts for advection, dispersion in three dimensions, sorption and first-order decay reactions. The EXPAR system was developed to assist users in the preparation of input data for ROKEY. The system is partitioned into two levels of operation. At the top level is a set of computer forms that contain groups of related parameters and corresponding entry fields. Users can either volunteer parameter values or request help from an assistance program. Elaboration programs that provide supporting information for each parameter and brief tutorials for the transport processes accounted for in the model are included. The set of assistance and elaboration programs represent the bottom level of EXPAR. The EXINS system was designed to assist users in planning the first stage of a monitoring strategy. The strategy is based mainly on the user's responses to a series of questions and checklists.

The results of an evaluation exercise demonstrated that EXPAR was partially unsuccessful at providing assistance but it was successful in terms of human engineering aspects. It is anticipated that one more development stage would be needed before the system can attain a high level of technical performance. Based on an informal assessment, it appears that the design of EXINS will be suitable for accomplishing its task. Therefore, continued development is warranted.

The results of the project have shown that expert systems techniques can be applied successfully to solve problems associated with the implementation of a transport model. The project has also demonstrated that these techniques could be successful at alleviating the complexity of planning a monitoring strategy. Given the expense of building transport models and conducting monitoring programs, the further application of expert systems techniques in these two areas merits attention. In addition, the feasibility of applying these techniques to solve problems in other areas of hydrogeology should be assessed.

ACKNOWLEDGEMENTS

I extend my greatest appreciation to my supervisor, Dr. F.W. Schwartz, for suggesting such an innovative research topic. His guidance and support throughout the project were invaluable. The comments and suggestions made by the members of my examining committee, Dr. K. Muehlenbachs, Dr. B. Nesbitt and Dr. P. Huck, are gratefully acknowledged.

The completion of the Expert ROKEY computer system required the assistance of many individuals. The contributions provided by my colleagues Ronald Collette, Allan Crowe, Alan Fryer, João Küpper, Bill Milne-Home and Rob Schincariol are sincerely appreciated. The original HOKEY code was provided by Dr. P. Domenico (Texas A&M). Values of organic carbon content were submitted by Dr. J. Barker (Waterloo). I would like to thank the students in the 1987 class of Geology 686 for their participation in the evaluation exercise. I am indebted most of all to the support, encouragement and assistance given by Dr. R.E. Jackson of Environment Canada.

Funding for the project was generously provided by the Alberta Environmental Research Trust Fund, the Department of Supplies and Services, and by Environment Canada.

Lastly, I would like to thank Gosia and Thomas for their patience and understanding.

TABLE OF CONTENTS

Chapter	Page
1. INTRODUCTION	1
2. OVERVIEW OF EXPERT SYSTEMS	3
2.1 Types of Knowledge	3
2.2 Traditional Expert Systems	4
2.2.1 Production Rules	4
2.2.2 Semantic Nets	9
2.2.3 Frames	11
2.2.4 User/System Interaction	15
2.3 Partitioned Expert Systems	17
2.3.1 The ONCOCIN Expert System	17
2.3.2 The ELAS Expert System	20
2.4 Why Build an Expert System?	23
2.5 Programming Considerations	26
3. PROBLEM SELECTION AND OBJECTIVES	28
3.1 The EXPAR Expert System	28
3.1.1 Domain and Objectives	28
3.1.2 Knowledge Engineering	29
3.1.2.1 Site-Related Information	29
3.1.2.2 Chemical Information	33
3.1.2.3 Output Control Data	33
3.1.2.4 Tutorial Information	34
3.2 The EXINS Expert System	34
3.2.1 Problem Domain	34
3.2.2 Objectives and Domain Knowledge	37
3.3 General Objectives	39
4. THE EXPERT ROKEY COMPUTER SYSTEM	41
4.1 System Overview	41
4.2 Description of ROKEY	41
4.3 The EXPAR Expert System	45
4.3.1 Description of the System	45

Chapter Page

4.3.2 The Top Level 46
4.3.3 The Bottom Level - Elaboration Programs 51
4.3.4 The Bottom Level - Assistance Programs 53
4.3.5 Design Considerations 59
4.3.6 Discussion 62

4.4 The EXINS Expert System 64

4.4.1 Description of the System 64
4.4.2 Discussion 70

5. SYSTEM EVALUATION 73

5.1 Introduction 73

5.2 Site Description 74

5.3 Description of the Exercise 79

5.4 Results and Discussion 81

6. CONCLUSIONS 88

REFERENCES 91

APPENDIX A References Cited in Expert ROKEY 94

APPENDIX B EXPAR Computer Forms 99

APPENDIX C Principal Tutorial Topics in Elaboration Programs . 109

APPENDIX D Summary of Contents of Assistance Programs . . . 112

APPENDIX E Strategy File (K.BAS) in EXINS 118

LIST OF TABLES

Table		Page
1	ROKEY Input Data Parameters	30
2	Strategy Topics in EXINS	38
3	Parameter Groups in EXPAR	50
4	Types of Sources Considered in EXINS	67
5	Data Sets for Evaluation Exercise.	80
6	Derivation of Parameter Values	85

LIST OF FIGURES

Figure		Page
1	General Types of Production Rules	6
2	Architecture of a Rule-Based System	7
3	Example of a Semantic Net	10
4	Example of a Frame-Based System	12
5	Stages in a Monitoring Program	35
6	Modules of the Expert ROKEY Computer System	42
7	Comparison of ROKEY and DPCT Results	44
8	Top Level Architecture of EXPAR	47
9	Computer Form Design	49
10	Bottom Level Architecture of EXPAR	52
11	Architecture of EXINS	65
12	Pieces in "Locations of Monitoring-Well Nests" Topic	69
13	Two Rules in EXINS	71
14	Location of the Gloucester Landfill	75
15	Hydrostratigraphic Units at the Gloucester Landfill	76
16	Actual and Simulated Plume Configurations	78

1. INTRODUCTION

One of the most important achievements in the field of artificial intelligence has been the development of innovative computer programs known as expert systems. Expert systems are designed to assist users in solving lengthy or difficult problems quickly and effectively. These systems have this capability because they contain the specific knowledge utilized by experts to solve a well-defined type of problem and operate by applying this expertise to a user's particular problem. Most systems include features that provide detailed descriptions of the problem-solving process and supporting information that will increase the user's understanding of the problem. Consequently, an expert system can become a valuable tool for enhancing a user's overall capabilities and productivity.

Since their inception in the early 1970's, many systems have been built in diverse areas such as chemistry, medicine, business and geology. The pioneering systems include DENDRAL, MYCIN and PROSPECTOR. DENDRAL interprets experimental data to determine the chemical structure of an unknown compound. The MYCIN system diagnoses infectious blood diseases and recommends treatment plans. PROSPECTOR assists geologists with the evaluation of sites for the presence of various types of ore deposits. More recent applications of expert systems in geology include ELAS and Dipmeter Advisor, both of which assist in the analysis of oil well log data. Prior to the completion of this project, expert systems in hydrogeology were not available.

This thesis examines the design and application of expert systems to problems in contaminant hydrogeology. The investigation involved the

development of a computer system that contains two expert systems. One system is designed to assist and advise a user in preparing a set of input data for a contaminant transport model, and tutor the user in the transport processes represented in the model. The second system recommends the first stage of a monitoring strategy.

Chapter 2 gives an overview of expert systems for those who are not familiar with the technology. The problems selected in this project for which the two expert systems were to solve are discussed in Chapter 3. A summary of the objectives for the expert systems and the computer system also is presented in this chapter. Chapter 4 provides detailed descriptions of the designs of the two expert systems. Due to the sizes of the systems, hydrogeological descriptions are limited to the presentation of representative examples. The results of an evaluation exercise in which a group of test subjects applied the computer system to an existing contamination site are presented and discussed in Chapter 5. The final chapter gives the concluding remarks for the project and discusses the relevance of the research to other problems in hydrogeology.

2. OVERVIEW OF EXPERT SYSTEMS

A brief discussion of expert systems is presented in this chapter for those readers who are unfamiliar with the technology. Numerous architectures, structures and styles of user/system interaction have evolved over the past decade. Therefore, the emphasis in this chapter is on those features that relate to this thesis project. Much of the following is based on information presented in the introductory texts written by Goodall (1985), Waterman (1986), and Weiss and Kulikowski (1984).

2.1 Types of Knowledge

Expert systems have the capability of assisting users in solving lengthy or difficult problems because they contain and allow the user to access a body of problem-specific knowledge. This knowledge is obtained either directly or indirectly from people who are experts in the problem domain. These systems do not operate by applying complex theories of reasoning to a given problem. Rather, it is the logical manipulation of the knowledge that leads to the solution to the problem.

An expert is a person capable of solving complex problems within a specific area. This ability comes from extensive experience and highly specialized knowledge. Building an expert system (a process referred to as knowledge engineering) involves the extraction and encoding of an expert's problem-solving knowledge so that others may have the same problem-solving ability. Generally, there are two types of knowledge involved:

(a) public information obtained from textbooks, articles in journals,

manuals, conventional computer programs, and data bases;

(b) private information gained from problems solved in the past.

The information in (b) is referred to as heuristic or, more simply, "rule-of-thumb" knowledge. Heuristic knowledge is commonly a collection of simple facts, general beliefs and hunches, and insights into the best procedures for solving a problem. It may be common to all individuals experienced in a domain or it may be unique to one individual. High quality heuristic knowledge is a rare commodity. Expert systems that have it and use it effectively can be of great value.

2.2 Traditional Expert Systems

The knowledge obtained from an expert is placed into an explicit, uniform structure that will permit the application of consistent methods of processing (Weiss and Kulikowski, 1984). Over the past ten years, three principal structures have been used in building expert systems: production rules, semantic nets and frames (Goodall, 1985; Waterman, 1986). I refer to those systems that employ one or a combination of these as being a traditional system. Descriptions of these structures, the ways in which they are implemented in the problem-solving process, and the style of user/system interaction commonly used are given in the following sections.

2.2.1 Production Rules

The production rule formalism is the most widely used structure for representing knowledge. Production rules are simply 'If-Then' conditional statements that are processed or "triggered" when any of the data gathered for a particular problem match the data in the 'If' part (i.e., the antecedent) of the rule. There are three types of

production rules that can be used in an expert system (Negoita, 1985). The general form and an example of each are shown in Figure 1. An inference rule consists of an antecedent which, when found to be true, leads to the assertion that the consequent ('Then' part) also must be true. When one of these rules is triggered, the information in the rule's consequent either replaces some of the information already gathered or is added to the collection. Premise-conclusion rules are similar but they are used to express intermediate and final conclusions reached during a session. Situation-action rules apply procedural knowledge. The triggering of these rules can lead to a variety of actions, such as the printing of text on the video screen, an interaction with a physical device (e.g., the on/off switch on an instrument), the entry into another system (e.g., an external data base) or a call to a programming procedure. In some instances, the user may provide information that is known to have a measure of uncertainty associated with it. To account for "reasoning under uncertainty" it is common for a system to employ some type of heuristic scoring function, such as a certainty factor. An example of a rule that uses certainty factors is shown in Figure 1d.

An expert system that employs production rules is called a production system or a rule-based system. The general architecture of these systems, based on Alty and Coombs (1984) and Weiss and Kulikowski (1984), is shown in Figure 2. The computer code is stored in the Program Control Unit, and includes the mechanisms that select and process the rules. The Knowledge Base contains the domain-specific knowledge obtained from the expert. Data for a particular problem often are obtained from the user through a questioning scheme. If requested

(a) Inference Rule

If: A is true and
B is true and
C is false

Then: X is true

eg. If: the chemical is Phenol and
the bacterium is Pseudomonas putida and
the bacterium is fully acclimated and
aerobic conditions exist

Then: the half-life may be in the order
of hours or a few days

(b) Premise - Conclusion Rule

If: A is true and
B is false

Then: Conclude X

eg. If: the mobility factor is 6 or greater and
the persistence factor is 4 and
the toxicity factor is 3

Then: emergency response actions should be
taken

(c) Situation - Action Rule

If: A is true and
B is true

Then: take a specified
action

eg. If: the dispersivity is known and
the velocity is known and
the effective diffusion coeff. is known

Then: call the procedure that calculates the
dispersion coefficient

(d) Conditional Inference Rule

If: A is known with a confidence of 0.9 and
B is known with a confidence of 0.4

Then: X is true with a confidence of 0.5

Figure 1 General Types of Production Rules

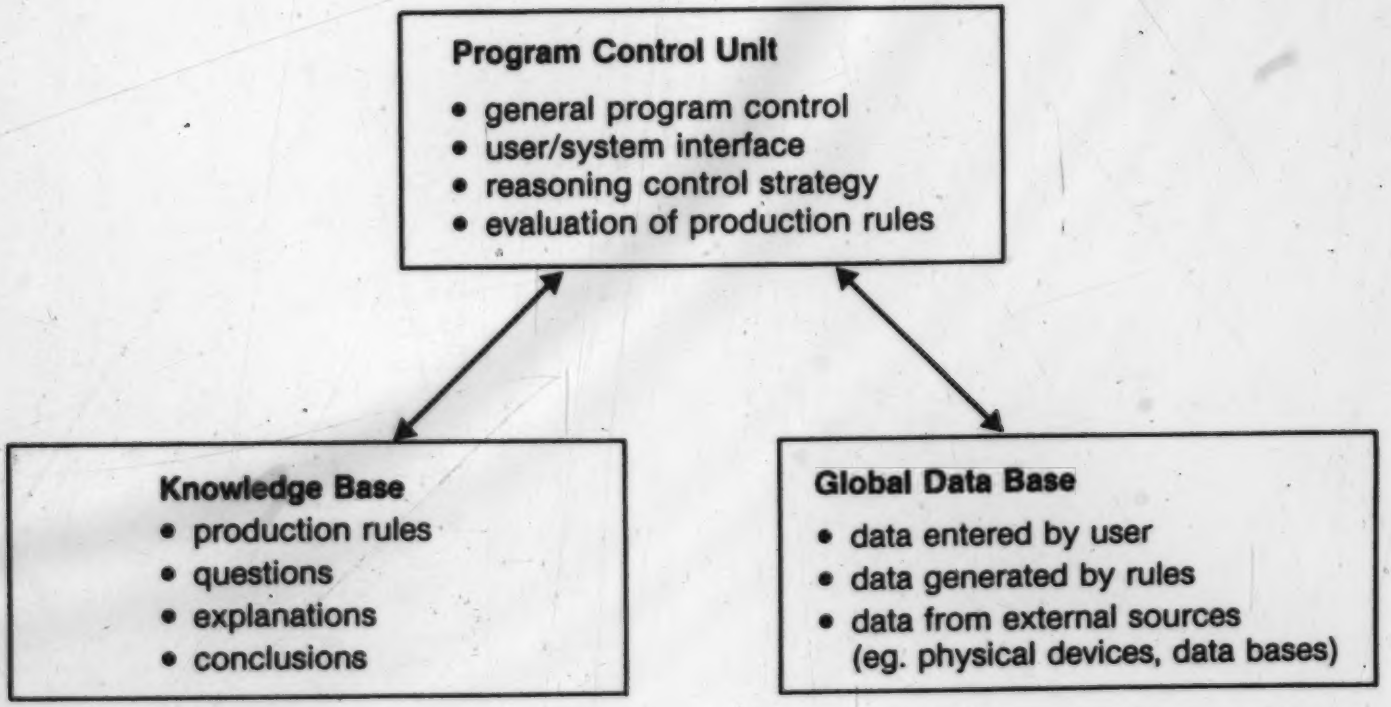


Figure 2 Architecture of a Rule-Based System

by the user, an explanation will be provided for each question. This feature of the system is discussed in Section 2.2.4. The Global Data Base (GDB), which is empty at the start of a session, stores the information gathered during the session.

A set of rules forms an inference network when it is interpreted in a session. Unlike binary decision trees, each rule in a network can connect to several other rules. At the start of a session, the user will be asked a question. The response is placed into the GDB. The mechanisms in the Program Control Unit then process the response by determining which rule applies and triggers it. In the case of an inference rule, this will result in the addition of data to the GDB. The rule set is then examined for another match between the data in the GDB and the antecedents of the rules. This procedure continues until a decision is made by the triggering of a premise-conclusion rule. If the process stops before a conclusion is reached, additional information will be sought by asking the user another question. The processing of rules in this manner is regarded as a form of deductive reasoning. The successive collection of problem-specific data causes the rules to trigger in a chain reaction that is analogous to a deductive reasoning scheme. The process described above, referred to as forward chaining, is one of several different methods of rule processing.

There are several important features associated with production rules. First, a significant amount of an expert's knowledge is inferential which makes inference rules a natural way of expressing this knowledge. Second, production rules are modular units of knowledge which can be modified with relative ease. Third, production rules enable knowledge to be represented explicitly and the contents of the

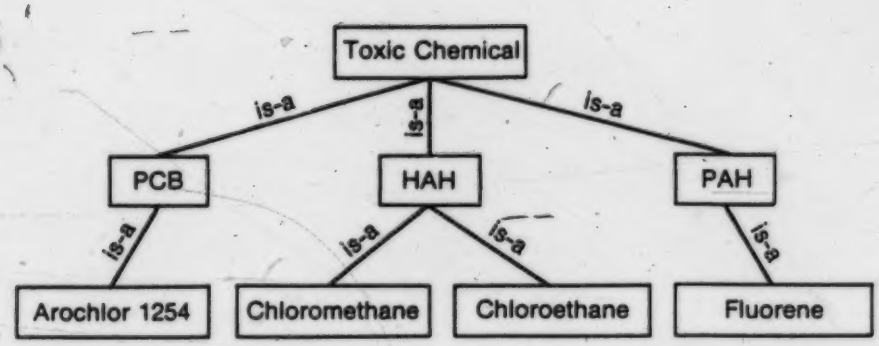
rules are readily identifiable in most computer languages. Consequently, experts not familiar with computer programming who wish to modify the system can scrutinize a rule set without difficulty. These features illustrate the key reasons why rule-based systems are popular.

2.2.2 Semantic Nets

Semantic nets are a popular means of representing non-rule knowledge. The concept of the semantic net is based on the idea that the human memory functions in part with associations between objects, concepts and events (Goodall, 1985; Waterman, 1986). A segment of a hypothetical net is illustrated in Figure 3a. The two main components are nodes (e.g., Chloromethane) that are linked by arcs (e.g., is-a). Each node is named. Other arcs commonly used are 'has', 'owns', 'needs' and 'is-part-of'.

The most important feature of a net is the processing power afforded by the 'is-a' link in building hierarchies of concepts. The nodes that are low in the hierarchy automatically inherit the properties of those higher up. Therefore, an expert system that makes use of the net in Figure 3a implicitly infers that 'Arochlor 1254' is both a 'PCB' and a 'Toxic Chemical'. It is because of this inheritance characteristic that semantic nets are used in rule-based systems to reduce the number of rules needed in the knowledge base. This concept is illustrated with the rule shown in Figure 3b and the network fragments given in Figure 3c. If the expert system presents the question "Is Arochlor 1254 present at the site?" and the user provides an affirmative response, the system will conclude that the facility is a hazardous waste site. The inference process was handled by the

(a)



(b) If: the compound is-a Toxic Chemical and
the facility overlies a shallow aquifer and
leakage has been detected

Then: the disposal facility is a hazardous
waste site

(c) Arochlor 1254 $\xrightarrow{\text{is-a}}$ PCB

PCB $\xrightarrow{\text{is-a}}$ Toxic Chemical

Figure 3 Example of a Semantic Net

semantic net and the conclusion was derived with the production rule.

In expert systems that combine rules and networks, the network fragments are stored in the knowledge base with the rules. The program control unit consists of two reasoning mechanisms, one that handles rule operations and another that controls net operations.

2.2.3 Frames

Objects and events that are stereotyped by a number of attributes are a common feature of many problem domains. A frame is a knowledge representation structure that is specifically designed to take advantage of this. A frame contains a group of related parameters called slots to which values are assigned. Associated with each slot are one or more procedural attachments that are utilized to control the use of the slot. The most important of these is the 'If-needed' procedure. This feature is used to assign slot values and operates by retrieving values from the global data base via the user. In addition to procedures, some slots have pieces of data attached to them, such as default values, plausible values, error values and measures of importance. A frame-based expert system consists of several frames that are organized into a hierarchy. The topmost frames represent general objects or events while the lower ones contain more-detailed information.

An example of a hypothetical system is illustrated in Figure 4. This system is designed to determine the hazard rating of a site and the recommended action (e.g., regulatory or remedial) that should be taken. The system was prepared on the basis of the SIA (Surface Impoundment Assessment) method as described by Canter (1985). Each of the frames describes an object that is stereotyped by several

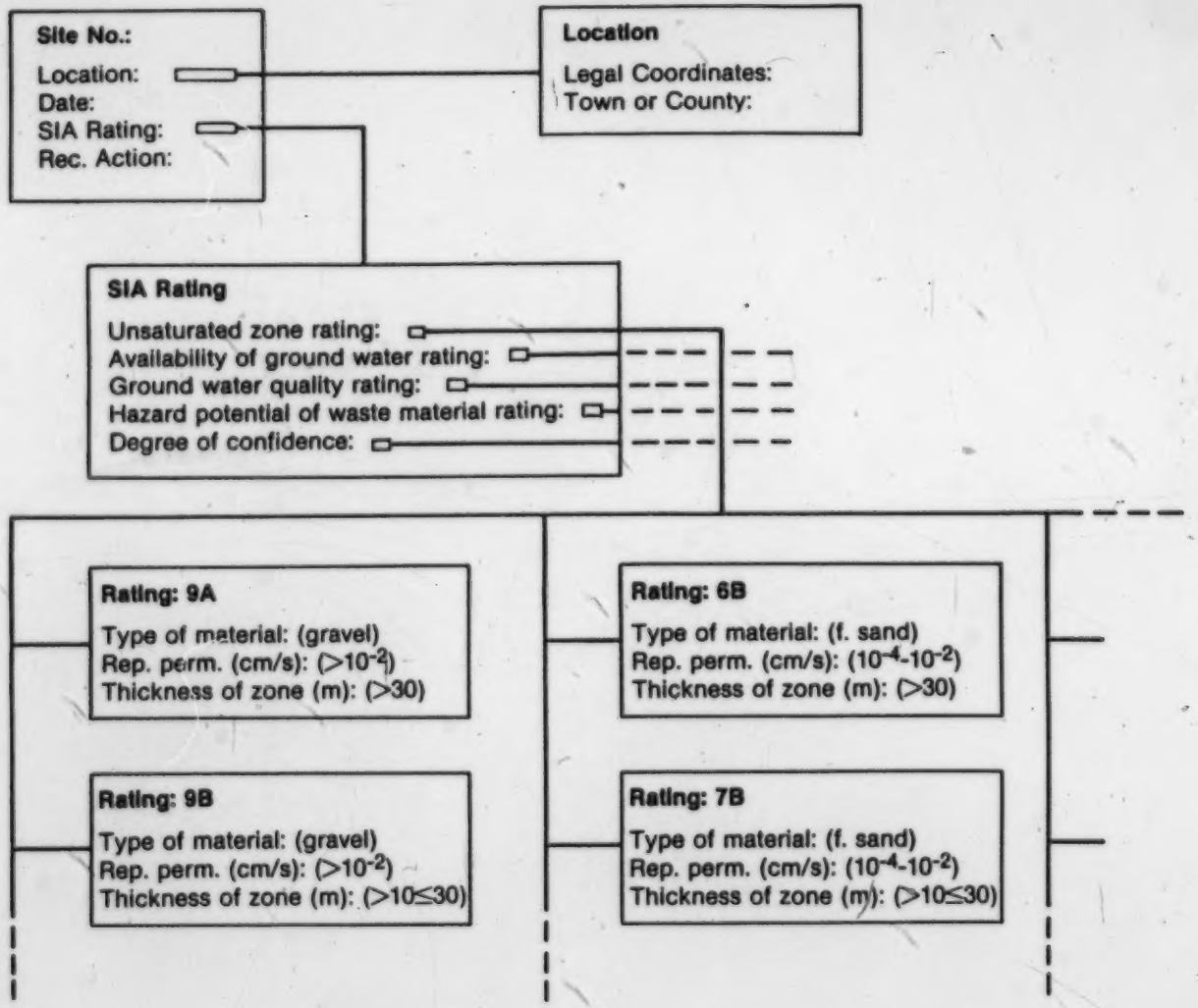


Figure 4 Example of a Frame-Based System

parameters. For example, the object 'SIA Rating' is stereotyped by five parameters. All five of these are called slots. The bracketed entries (e.g., gravel) in the Rating frames represent plausible values.

The user's task in working with the example system is to assign slot values. Some values can be assigned easily, such as 'Date', while others may require detailed analysis. As an example, the user would be asked whether an SIA rating is available for a particular site. A negative response would activate the SIA Rating frame and questioning would be focussed on the derivation of values for the five slots associated with the frame. A negative response to the question "Do you have a rating for the unsaturated zone?" would result in questioning that would attempt to identify the type of material at the site, a representative permeability value and the thickness of the zone. Responses of "fine sand", "0.001" and "40" would produce an unsaturated zone rating of 6B. The processing of the frames would continue in this way until an SIA Rating is determined. The system would provide a default value for a slot if the user was unable to assign a plausible one. Also, all of the entries would be checked for consistency using the error values attached to each slot. It should be noted that frames are described with programming notation and are not displayed on the video screen.

The representation and organization of knowledge in a frame-based system is characterized by the following features:

(a) a frame can inherit the properties of other frames;

(b) frames provide a means of grouping together similar objects, events and attributes; this knowledge of association assists in making all of the domain knowledge more explicit compared to a rule-based system;

(c) control knowledge is represented by the procedural attachments associated with each slot;

(d) inferential reasoning occurs implicitly in a frame system as case-specific data are assigned to each slot.

Similar to expert systems that use semantic nets, a frame-based system is very useful for representing knowledge in domains that are characterized by well-established taxonomies.

A frame-based system can be modified in a number of ways to represent and apply domain knowledge more efficiently and effectively. One example is the CENTAUR system for the diagnosis of pulmonary diseases (Aikins; 1983,1984) that uses production rules as added procedural attachments. Some of the slot operations are performed with situation-action rules and values for the slots can be derived with inference rules. This design is accomplished through the use of 'pointers' that link the rules in the rule set with the slot to which they apply. There are two main advantages of combining frames and production rules. First, inferential and associational knowledge can be expressed together with the frames conducting the general reasoning and the rules performing the fine-grained reasoning. Second, the rule pointers designated for each slot explicitly show which rules are associated with the slot. In a system consisting of hundreds or even thousands of rules, this latter feature can be important when rules are added, deleted or modified.

A frame-based system, with or without production rules, is similar in architecture to a rule-based system (Fig. 2). The domain knowledge is stored in the knowledge base and the mechanisms for operating the system are contained in the program control unit.

2.2.4 User/System Interaction

Most of the traditional systems communicate with the user by engaging him or her in a type of dialogue. This interaction simulates a conversation between an expert and the person with the problem. The system will request information by presenting a series of questions on the video screen, to which the user will respond on the keyboard with a truth value (yes, no, unknown), a number, a word or a string of words. Intermediate conclusions (if any) and the solution to the problem will be displayed on the screen in the form of words or sentences. This question-answering (QA) interaction is still used frequently. However, other forms of communication may be employed, such as having the user select an item from a checklist or "menu".

The majority of systems contain an explanation facility. The meaning of some of the questions presented may not be clear or the user may not understand why a question is being asked. By responding with a "?" or in a similar way, the system will provide an explanation or a rephrasing of the question. In many cases, the system also will explain the chain of reasoning that is currently being followed. After the solution to the problem has been given, a report may be requested by the user that explains how the system arrived at the solution. This feature is important because it educates the user and increases his or her confidence in the reliability of the solution. An explanation facility can be just as important as the solution to the problem in some systems.

The question-answering scheme is effective at communicating with the user but it has several major drawbacks from a human engineering perspective. First, for systems that require a large amount of data,

the user could be faced with the tedious task of answering dozens of questions. Second, users can not readily volunteer information at any time. Instead, they have to wait until the appropriate question is asked. Third, the user can not return to a previously answered question in order to change the answer. Therefore, the session may have to be halted and a new one begun. Fourth, for systems that require data for the same set of parameters (referred to as static parameters), users will encounter the same set of questions with each session. This could eventually make the use of the system somewhat monotonous. Fifth, users can not retrieve a set of their responses made in an earlier session, change the value given to one of them and then obtain a new, possibly different solution. Again, the user would have to initiate a new session and reanswer the questions. Finally, most systems allow the user to enter "bookkeeping" information (e.g., the date and time of the session) that may be useful for later reference but is not required for the derivation of a solution. In a QA scheme all questions must be answered, including those related to bookkeeping, in order for the session to proceed. The inability to avoid nonessential questions can be a source of irritation for some users.

An effective way of minimizing these problems is to gather some of the information from the user with some type of computer template or form. In essence, a computer form is an explicit version of a frame in that related parameters are displayed on the video screen along with entry fields into which the user or the system can enter data. The introduction of these forms in expert systems has lead to the development of new ways of representing, organizing and controlling domain knowledge. One example is in what I refer to as partitioned

expert systems.

2.3 Partitioned Expert Systems

A partitioned system consists of at least two distinct programs. One program is devoted to the acquisition of data for the static parameters. The second is used for gathering values for dynamic parameters and for processing of all of the information. Although separate, the programs usually run in parallel and communicate interactively with each other. Two partitioned systems that had a major influence on one of the expert systems developed in this project are ONCOCIN, an expert system used in medicine, and ELAS, a system used in petroleum geology. A brief description of both systems is given in the following sections to illustrate how computer forms can be used and how they have lead to the development of new knowledge representation structures. Detailed descriptions of ONCOCIN and ELAS can be found in Buchanan and Shortliffe (1984), and in Weiss and Kulikowski (1984), respectively.

2.3.1 The ONCOCIN Expert System

ONCOCIN is a descendent of MYCIN, the most well-known expert system in medicine. It was designed to assist physicians in selecting a chemotherapy treatment plan for cancer patients. Traditionally, physicians would enter a patient's data (e.g., lab test results) on a paper form during each visit. A treatment plan then would be devised on the basis of the information for the current and past visits. Developing a plan is a complex task because it has to be selected from a large number of possible plans (30 to 60 depending on the treatment institute) with each one described in a detailed 40- to 60-page document.

ONCOCIN was built to help physicians with this task. The system is comprised of two programs: an interface program (the Interviewer) for the entry of data and a reasoning program (the Reasoner) which determines an appropriate treatment plan. In a session with the system, the physician uses the Interviewer to enter the patient's data for the current visit and to retrieve data that were derived from previous ones. All of the data is passed on to the Reasoner which uses a set of production rules to formulate a treatment plan. The plan is comprised mainly of recommended dosages of several drugs that are used in the therapy. Once formulated, the plan is passed back to the Interviewer which, in turn, displays it to the physician. The physician can either accept the plan or make changes to the dosages as he or she deems appropriate.

The Interviewer utilizes a video display that is divided into four regions. The largest one, a computer form that resembles the physician's paper form, can have up to eight columns at any one time. The left-most column has a list of static parameters (e.g., white cell count, platelet count) including the names of the drugs for which dosages can be recommended. Up to six columns to the right of the list contain parameter values that were derived from previous sessions. The right-most column contains the values entered by the physician for the current session; after the Reasoner has processed the data, the recommended drug dosages are displayed at the bottom of this column. Of the three other regions on the video display, two are used for presenting messages and explanations, while the other indicates the special functions associated with several keys positioned at the top of the keyboard.

Information is entered with a customized keypad that includes a 10-key number set, 'YES' and 'NO' keys, and cursor control keys for moving to individual entry fields. As the physician enters data, the message region displays the range of expected values for the parameter with which the cursor is aligned. After the Reasoner has transferred the recommended treatment plan back to the screen, a brief justification for the recommended drug dosages will be shown in the explanation region.

The domain knowledge in ONCOCIN is expressed with four structures in addition to the set of production rules. These structures are: contexts, parameters, data blocks and control blocks. Contexts represent the concepts of the domain about which the system needs static knowledge. The parameters consist of the attributes of patients, drugs, tests, etc. The total set of parameters includes the static parameters listed on the display and those that are internal in the Reasoner. Each piece of information accumulated during a session is represented as the value of a parameter. In most cases, values are either entered by the user, determined by definition in the current context or concluded using the set of production rules. If these methods fail, rules that provide a default value are triggered. The data blocks define logical groupings of related data. The system treats data blocks as units when obtaining values from the user, and during the storage and retrieval of data in a patient's file. Control blocks contain the procedures the system is to follow when a task is specified by the user. These may include the retrieval of data blocks, the determination of parameter values, the invocation of other control blocks and calls to special-purpose functions.

The entry of information using the computer form represents an important feature in terms of the acceptance of the system by the users. Physicians are familiar with the paper form already so a computer form makes the interaction more comfortable. Completing the form is fast, easy and simple to learn. Users have the freedom of entering only the information they feel is pertinent, leaving some entry fields blank if they wish. The data can be entered in any order, and current and old data can be changed at any time. Most importantly, computer forms provide a means of allowing the user to volunteer information.

2.3.2 The ELAS Expert System

ELAS is an expert system that assists users in the analysis of oil well log data. The system consists of two integrated FORTRAN programs: Amoco's interactive program called INLAN that was developed for well-log data analysis, and an expert system that coordinates and provides advice on the use of INLAN. In developing ELAS, the main objective was to demonstrate how the knowledge and reasoning methods of an expert log analyst could be combined with existing software to produce a system that would be of use to a wider variety of expert and nonexpert users.

The main task of a log analyst is to determine the presence or absence of hydrocarbons in one or more zones of a borehole based on his interpretation of a set of logs. The interpretation is made after the values for several parameters have been derived and numerous tasks have been completed. The parameters include the porosity of the rock, the saturation of the formation water (Sw) and the resistivity of the water (Rw). The tasks may include the preparation of 1/Sw histogram plots and

neutron-density crossplots. Prior to the development of ELAS, the analyst would derive the values and perform the tasks using the components of INLAN directly. The sessions were experimental in nature where values would be changed frequently and tasks rerun in order to assess various outcomes and interpretations. The processing of the logs in this way required extensive experience in the analysis of log data plus considerable skill in operating INLAN. The motivation behind the development of ELAS was to give nonexperts the ability to correctly interpret log data and to make INLAN easier to use.

In ELAS, the use of the components of INLAN are coordinated through a computer form, called the master panel, which is displayed on a video screen. The panel is divided into three regions much like the video display in ONCOCIN. The top region contains several parameters that provide a general description of the borehole (bookkeeping information). The region below this is divided into six columns with the left-most column consisting of a list of the key parameters and tasks that are crucial in a log analysis session. The remaining columns contain entry fields for parameter values for up to five target zones of the borehole. The third region contains a list of the commands used in conducting a session.

The user's objective in a session is to derive values for the key parameters listed in the master panel. Once a set of values has been obtained for a zone, the system will process it with a set of production rules that will provide an interpretation. The user can either volunteer parameter values or request assistance. If the latter is chosen, a subpanel will be displayed that lists the components of INLAN recommended for use at that point in the session. The entries on

the list are placed there by a set of production rules. After the user makes a selection, a component of INLAN will calculate a value and place it in the appropriate entry field in the master panel. The selection and execution of tasks is accomplished in the same way.

Most of the knowledge in ELAS is represented in the production rules. However, the list of parameters and tasks in the master panel also is an expression of knowledge. The subpanels from which the user makes selections can be regarded as control blocks as defined in ONCOCIN. Similarly, the sets of data determined for the target zones are knowledge structures (i.e., data blocks).

ELAS is a multi-level system where the expert systems components are surface level models of reasoning and the mathematical components of INLAN are deeper levels. The notion of multi-level systems was first expressed in an article by Hart (1982) in which he described a hypothetical system for a petroleum reservoir simulator. In this system, the guidance for analysing a reservoir is provided by a rule-based system (surface level model). The detailed analysis is performed by a set of partial differential equations (deep level model). The equations could represent various components of the simulator, such as Darcy's Law and the Equation of Continuity. The appeal of developing an actual system comes from Hart's recognition that the costs of building simulators are substantial and that their use is limited by the need for a large number of parameter values. Hart (p.13) concluded that "well-codified science and engineering disciplines offer perhaps the best domains in which to explore multi-level systems". The existing codes for contaminant transport models are very similar to a reservoir simulator in terms of

importance, complexity, cost and data requirements. Therefore, it is possible that the development of multi-level systems that include these codes could become valuable tools in contaminant hydrogeology.

2.4 Why Build an Expert System?

The most important benefit that can result from building an expert system is that users will be able to solve problems that were previously too complex or too lengthy. Therefore, the use of an expert system can enhance the capabilities and productivity of the user. Other reasons why expert systems are potentially useful are detailed by Waterman (1986), and by Weiss and Kulikowski (1984). The list includes:

(a) Expert systems provide a means of replicating and disseminating rare and costly expertise. This feature could be important to organizations for the training and educating of new personnel, particularly in situations where the key people who have the expertise leave.

(b) Expert systems help structure and organize the systematic components of problem-solving in a specific domain. This use is important for the problem-solving methods that are informally specified because an expert system requires the formalization of all knowledge.

(c) Expert systems provide a means of expressing heuristic knowledge. Once incorporated into a system, this knowledge can be tested and reproduced.

(d) The expertise of several people can be combined in an expert system. (This is important in contaminant hydrogeology where the solutions to many problems require experts from a variety of disciplines.)

(e) Expert systems can be powerful tools for predictive modeling. They can provide answers for a given situation and then show how these would change for new situations. This allows the user to evaluate the potential effect of new data and understand their relationships to the solution. Consequently, an expert system can simplify and improve the quality of decision making.

(f) Multi-level systems make existing software more accessible to a larger number of users, and increase the speed and effectiveness with which the software can be used.

(g) Expert systems can be useful to those already proficient in the problem domain by increasing the speed at which they solve problems or by providing a second opinion. In the case of the latter, a mismatch between the system's answer and the answer obtained by the user alone will force the user to check where and why the mismatch occurred.

The traditional approach for disseminating expertise and problem-solving knowledge is through the use of specialized manuals. Several manuals are available that describe methodologies for dealing with problems in contaminant hydrogeology. Examples include the manual that describes the DRASTIC system for evaluating ground water pollution potential (Aller et.al., 1985) and a manual that outlines a pollution assessment method for emergency response conditions (Donigian et.al., 1983). Although manuals can be used effectively to solve a problem, they have several shortcomings that are avoided in expert systems, as indicated below:

(a) A considerable amount of effort often is required to learn how the manual and the problem-solving methods are to be employed.

This effort may be repeated every time the manual is needed if several months pass between usages. Expert systems guide the user through the session. In most cases, all that is needed is an understanding of the general purpose and operation of a system.

(b) Most manuals are large and contain a vast amount of detailed information. Consequently, the use of the manual may be time consuming and there is the risk that important information may be missed. Expert systems also can be large and they typically contain a great amount of detailed information. However, because computers can search and manipulate a large amount of information rapidly, sessions with a system can be relatively short. Also, if a system is well-designed, no details will be overlooked.

(c) Manuals usually require updating, which is difficult to do on a frequent basis. The structures used in expert systems are designed to facilitate easy and quick updating.

(d) For manuals that address a large and complex problem, the authors are able to present only a standardized approach to problem solving. Many of the nuances that may have a major bearing on the solution may be left out. By using a computer, all of the information related to a problem can be incorporated into an expert system.

(e) It is difficult to express the dynamic, systematic component of a problem-solving process in a manual. Some authors try to present this in large fold-out flow diagrams but these work only for relatively small problems. Flow diagrams for larger problems quickly become unwieldy. The structures used in expert systems are designed for representing all of the different paths that can be

taken in reaching a solution. Large flow diagrams are expressed implicitly within the system.

Expert systems take longer to build but the additional time and cost may easily be offset by the benefits to be realized.

2.5 Programming Considerations

Expert systems can be built using specialized knowledge engineering languages or they can be built "from scratch" with a standard programming language. Knowledge engineering languages are computer programs designed solely for the construction of expert systems. They consist of a program control unit and various programming tools for developing a knowledge base. Many of these languages (e.g., KAS, EMYCIN, EXPERT) are simply existing expert systems but with the domain-specific knowledge taken away. It is because of this that they are usually referred to as "skeletal systems" or expert systems "shells". Expert systems can be developed quickly and easily with a shell program. However, due to their lack of generality and flexibility, they can be used only for a restricted class of problems (Waterman, 1986). If the problem at hand does not fit the characteristics of the program exactly, it can not be used. Also, shell programs tend to be useful only for the development of small, simple systems. Any application with even a moderate degree of sophistocation soon extends beyond of capabilities of existing shell systems (Silverman, 1986).

The alternative is to build the expert system with a regular programming language. Most expert systems operate solely on the manipulation of symbols and, as such, are written in either LISP or PROLOG. However, other languages such as C, FORTRAN and PASCAL can be

used. The advantage to using a standard language is the developer has the flexibility to create a system that is tailored exactly to the requirements of the problem and to the perceived needs of the users.

3. PROBLEM SELECTION AND OBJECTIVES

The computer system developed in this project was designed to perform two tasks. The first one is to assist users in estimating the distribution of a chemical in a ground water flow system. The second is to recommend the first stage of a monitoring strategy. The computer system, called Expert ROKEY, consists of three main components, (1) a contaminant transport model, (2) an expert system that is used to prepare a set of input data for the model, and (3) an expert system that recommends a monitoring strategy. The names given to these components are ROKEY, EXPAR and EXINS, respectively.

The following sections outline the problem domains, objectives and knowledge engineering aspects of EXPAR and EXINS. The chapter concludes with a statement of the objectives defined for the entire computer system.

3.1 The EXPAR Expert System

3.1.1 Domain and Objectives

Predicting the current and future distributions of contaminants is an important task undertaken by hydrogeologists. This problem often may require the use of some type of contaminant transport model. While numerous models have been developed and are readily available, many hydrogeologists are not able to use them. For some, this inability stems from their unfamiliarity with models and computers in general. Others may be adept at working with computers but are not able to apply models effectively due to a variety of problems. For example, many models are sophistocated and their operation can be complex. Consequently, the user may need several years of modeling experience in

order to operate a model in the way it was intended (Schwartz et.al., 1985). Another problem is related to a model's input data requirements. There is little doubt that assessing the important parameters and assigning representative values is a formidable task (Domenico and Robbins, 1985). In most cases the collection of problem-specific data involves lengthy field programs, and for some parameters such as source size and loading, data simply may not be obtainable. A third problem also is associated with the input data. Some models require the construction of a large data set which can be a long and arduous task. In many instances, errors in data entry can occur and remain unnoticed. A fourth problem is related to the fact that the transport of chemicals in the subsurface is governed by many physical, chemical and biological processes. The proper use of a model requires an understanding of these. However, the processes are complex and interest in problems of ground water contamination has grown only recently. Consequently, many hydrogeologists have little or no understanding of the processes represented in the models that they use.

The main objective in designing EXPAR was to solve or at least minimize the problems outlined above through the use of expert systems techniques. Specifically, the system was to (1) advise and assist a user in preparing a set of input data for the ROKEY model, and (2) tutor the user in the transport processes represented in the model. In terms of expert systems research, a secondary objective was to explore the multi-level approach to system design as described in Chapter 2.

3.1.2 Knowledge Engineering

3.1.2.1 Site-Related Information

The input data parameters for ROKEY are shown in Table 1. Values

Table 1 ROKEY Input Data Parameters

Parameter Group	Parameters
Stratum	v - ground water velocity (advection rate) dx, dy, dz - dispersivity tensors in the x, y and z Cartesian coordinate directions D* - effective diffusion coefficient n - total porosity foc - organic carbon content bden - bulk density DWT - depth to water table DBB - depth to base of stratum
Source	YD - source size in y direction DTS, DBS - depths to top and bottom of source Co - initial concentration at the source Co/t - changes in concentration with time at the source
Chemical	Koc - organic carbon/water partition coeff. t1/2 - decay half-life for abiotic hydrolysis or biotransformation
Simulation	t - duration of simulation period
Output Control	XINT - X interval of grid YINT - Y interval of grid YCO - Y coordinate of XZ plane ZCO - Z coordinate of XY plane

for the stratum and source parameters ideally will be determined from information that is specific to the site under investigation. To assign values, experts typically gather and analyze all available information for the site. If the information contains a plausible value for one of the parameters it will likely be used in the model. If several plausible values are available that were derived by one method, the expert may select the highest or lowest one, or calculate an average value. This assignment generally is dictated by the objectives of the study, such as a best case/worst case analysis of the contamination problem. Where several values are available that were determined by more than one method, the selection will be restricted to the method that the expert believes yielded the most accurate results. In some cases the available data will require processing in order to derive a value. The expert may have several processing methods available and will use the method that he prefers or is best-suited for the data.

In virtually all studies, there will not be enough high-quality data available from which values for all of the parameters can be derived. Two main reasons are:

(a) the data are not available and the acquisition could involve a lengthy and expensive field program;

(b) the data are not available and it is not possible to acquire them.

A typical example of (b) is historical data on the types and volumes of chemicals disposed of at an industrial landfill.

To cope with these problems, experts are forced to use heuristic methods. Some of these include (1) estimating a value on the basis of related information, (2) choosing a value that was used at a similar site, and (3) using a default value. Default values usually are used

either as a last resort or when a "ball-park" estimate is suitable for the current needs of the study. In some cases the expert may have data available with which a plausible value can be calculated but will choose a default value rather than process the data. For example, the expert may decide to use a default value for porosity rather than perform the lengthy task of analysing grain-size distribution data.

Based on the above, the knowledge requirements of EXPAR would include:

(a) plausible and default parameter values for numerous hydrogeological conditions;

(b) rationale for choosing one value from several that were derived by the same method;

(c) problem-solving approaches, such as best case/worst case;

(d) methods to derive values directly and indirectly, and the associated rankings from most-preferred to least-preferred;

(e) listings of values determined from similar studies.

While some of this knowledge has been expressed formally, it is known that much of it has not. For example, methods used to determine values for rates of advective transport are discussed in many textbooks. However, a clear indication of which method is preferred over another is often not given. Preferences depend on several factors and, as such, the selection of a method will be based on the investigator's heuristic judgment.

The first task undertaken in building EXPAR was to collect and formalize all of the knowledge outlined above. The second task was to extract from the literature the reasoning strategies that could be employed in implementing the knowledge. Public sources of information

(textbooks, journals, manuals, etc.) were reviewed. The sources from which information was taken are included in Appendix A. Although this search provided most of the information used to complete the first task, few reasoning strategies could be found. Thus, part of this study involved the development of my own heuristic procedures.

3.1.2.2 Chemical Information

EXPAR is designed to provide input data for modeling the distributions of organic chemicals. The specific types of chemical data required were (1) octanol/water partition coefficients, (2) solubilities, (3) half-life values for abiotic hydrolysis and biotransformation reactions, and (4) maximum acceptable concentration limits. Because the number of organic chemicals associated with ground water contamination is in the thousands, the acquisition of data was restricted to the organic chemicals given on the U.S. EPA's 129 Priority Pollutant list. All of the data for the first three types was obtained from Callahan et.al. (1979) while concentration limits were obtained from the Federal Register (November 28, 1980). The data are included in two data bases. These two data bases can be expanded easily to accommodate the data for any number of additional chemicals.

3.1.2.3 Output Control Data

The output control data required by ROKEY are few. Users only need to be familiar with the documentation in the operations manual (McClymont and Schwartz, 1987) to assign values. However, it was anticipated that some users would be uncertain of the three-dimensional grid arrangement employed by the model. Therefore, assistance in selecting values for the parameters associated with the grid would be needed.

3.1.2.4 Tutorial Information

The effective application of the ROKEY transport model requires that users are knowledgeable about transport processes. This understanding must extend to the parameters which mathematically characterize these processes. Therefore, EXPAR would have to include short tutorials for the processes of advection, dispersion, sorption and transformation. Also required would be information for each parameter such as operational definitions, explanations as to why the parameter is part of the data set, descriptions of how the parameter is measured in the field or lab, discussions of related processes and theories, and lists of references.

3.2 The EXINS Expert System

3.2.1 Problem Domain

Another important task undertaken by a contaminant hydrogeologist is to plan and implement a ground water monitoring program. Monitoring programs are comprised of several stages, as indicated in Figure 5. The main objective is to assess ground water quality by obtaining and analysing quantitative information on the physical, chemical and biological characteristics of the water and the transporting medium. Once the assessment has been completed, the program usually terminates with some form of evaluation, such as whether remedial actions should be taken.

The most critical stage that may determine the overall success of a program is the installation of a sampling network and the collection of ground water samples. Considerations in carrying out these activities number in the dozens and include:

(a) assessing of the information needs of the program and establising

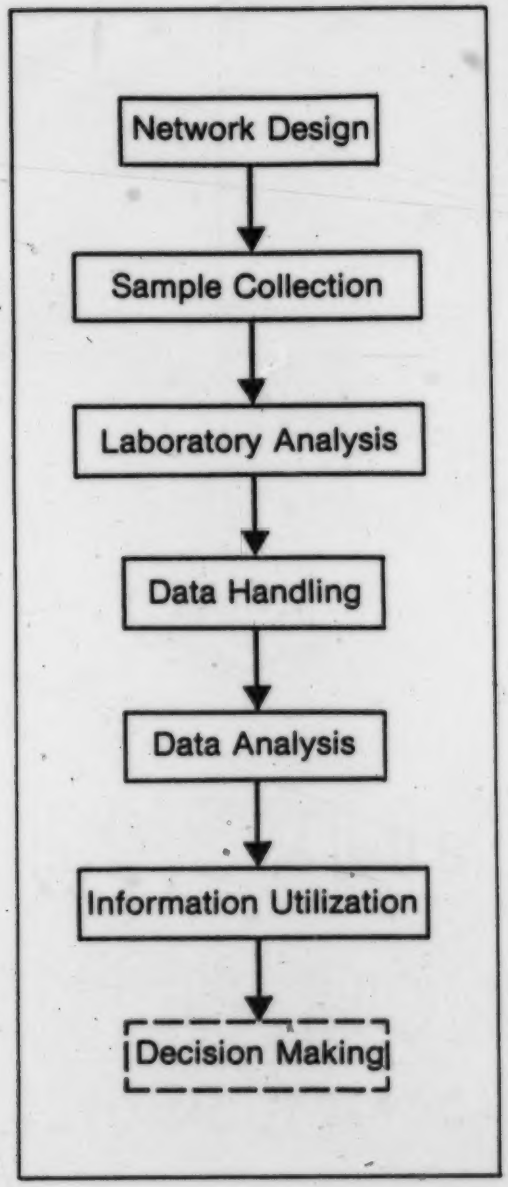


Figure 5 Stages in a Monitoring Program
(adapted from Sanders et.al., 1983)

quality assurance and quality control criteria;

(b) evaluating the hydrogeological setting;

(c) determining the types of wells needed for the network, and the number, locations and completion depths of each type;

(d) selecting drilling methods;

(e) selecting materials for construction of the wells, and designing construction and installation protocols;

(f) determining which chemical constituents are to be analysed and how frequently;

(g) establishing sampling and analytical protocols which might include deciding on the most appropriate sampling devices and materials, and the acceptable methods for sample collection, handling and analysis.

Preparing a strategy that accounts for all of the components is, with few exceptions, a difficult task. Further, a variety of problems can arise. For example, complexities likely will exist in the hydraulic, geochemical and geobiological conditions at a site. More than one contamination source may be present with each contributing a large number of different compounds at variable rates. Sampling for some organic compounds is difficult due to the possible volatilization of the compounds from samples, or the absorption of the compounds to the sampling materials. Added to these problems is the difficulty in finding commonly agreed upon procedures. Considerable controversy exists among laboratories, regulatory policies, and researchers concerning proper well construction techniques, sampling techniques and appropriate procedures for preserving the original chemical character of the samples (Barcelona et.al., 1984).

Consequently, the planning of a successful (and cost effective) monitoring program can require considerable expertise. The designer must be knowledgeable in hydrogeology, organic and inorganic chemistry and microbiology. Furthermore, he or she must be well-acquainted with government policy and regulatory requirements. Because no one person can be an expert in all of these areas, there is a need to develop a means of assisting designers, beginning with the development of monitoring design procedures that can be regarded as valid by most designers (Sanders et.al., 1983). To date, several researchers have produced texts and manuals that provide guidelines rather than widely agreed upon procedures. Hard-and-fast rules in such a complex area are difficult to develop (Sanders et.al., 1983), making the planning of a monitoring program more of an art than a science.

3.2.2 Objectives and Domain Knowledge

The main objective in developing EXINS was to investigate the potential for applying expert systems techniques to the preparation of monitoring program guidelines. The research was to be centered on the construction of a demonstration prototype that would make recommendations for 12 topics (Table 2) in the first stage of a monitoring program. The strategy would be designed for the site previously modeled by ROKEY. A demonstration prototype is a system that solves a portion of the problem, suggesting whether or not the approach is viable and system development is achievable (Waterman, 1986). Therefore, an assessment of the approach rather than the testing of the system is the final task undertaken at this stage of system development.

Most of the information incorporated into EXINS for the 12 topics

Table 2 Strategy Topics in EXINS

Topic No.	Topic
1	Types of Monitoring Wells
2	Locations of Monitoring-Well Nests
3	Completion Depths for Monitoring Wells
4	Monitoring Well Materials
5	Drilling Methods
6	Well Completion Methods
7	Ground Water Sampling Devices
8	Sampling Equipment Materials
9	Storage Containers
10	Sampling Frequency
11	Analytical Program
12	Sample Volumes and Sample Preservation

was obtained from public sources with minor contributions from my own heuristic judgments. The principal sources were "Procedures Manual for Ground Water Monitoring at Solid Waste Disposal Facilities" (Fenn et.al., 1977) and "A Guide to the Selection of Materials for Monitoring Well Construction and Ground-Water Sampling" (Barcelona et.al., 1984). Supplementary information was obtained from Cherry et.al. (1983) and Pettyjohn et.al. (1981).

3.3 General Objectives

In terms of knowledge content and human engineering considerations, the computer system (Expert ROKEY) was to be designed so that:

(a) it would be of use to a large number of hydrogeologists, with skills ranging from novice to expert;

(b) a minimal amount of skill in using a computer would be required, thereby making the system useful to a broad group of potential users;

(c) a minimal amount of time would be needed to learn how to use it effectively;

(d) disruptions would not occur due to simple user error.

In assessing the information requirements for the expert systems it was necessary to define a minimum level of knowledge in hydrogeology to be expected of the users. After considering system size criteria and deciding that the users should be professional hydrogeologists, the systems were designed for users with an educational background in hydrogeology plus a minimum of one or two years of relevant experience. Those who did not meet this level could have difficulties but the computer system could be used as an educational tool.

The computer system was to be developed as a practical instrument that could be applied to a variety of tasks. Several features were included in EXPAR that would make the computer system particularly useful for rapidly assessing the hazard posed by a leak or spill of a chemical in an emergency response situation, predicting the distributions of chemicals being discharged from municipal and industrial sources, and quickly screening proposed sites for waste disposal facilities.

4. DESCRIPTION OF EXPERT ROKEY

4.1 System Overview

The Expert ROKEY computer system consists of five modules (Fig. 6). The Introduction module contains five text files that provide (1) an introduction to the system, (2) information that orients the user to some of the tasks the system can be applied to, (3) a description of the case-specific information that would be useful to have on-hand for completing a session, (4) a brief description of the ROKEY model, and (5) a discussion of modeling strategies that may be employed. The Instructions module consists of one text file that provides a set of instructions for operating the system. The remaining modules (EXPAR, ROKEY and EXINS) are the focus of this chapter.

All of the modules in the system are integrated into one computer program. Therefore, conducting a session requires only one program call command. The system was designed to operate on a standard microcomputer with a hard disk. The total storage required is approximately 1.2 megabytes.

4.2 Description of ROKEY

ROKEY is an analytical transport model that calculates the concentration distribution of a chemical in a saturated porous medium. The model is based on the Extended Pulse Approximation model developed by Domenico and Robbins (1985) but was extended to account for first-order transformation reactions. The resulting solution is:

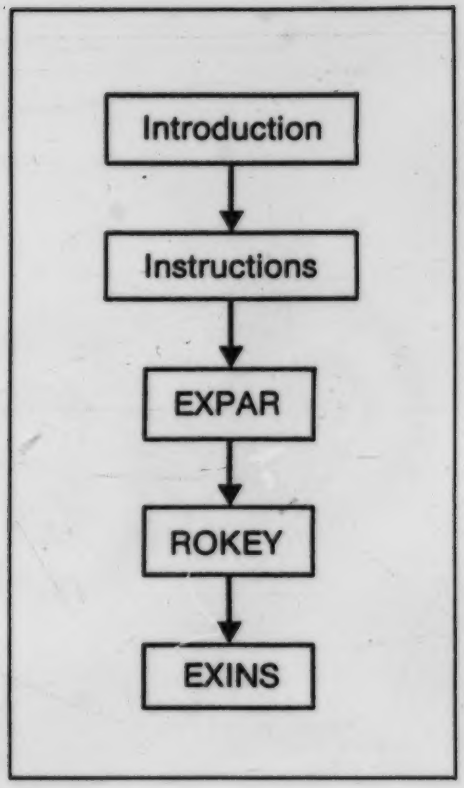


Figure 6 Modules of the Expert ROKEY Computer System

$$C(x,y,z,t) = (Co/8) * \exp[(v-u)*x/2Dx] * \{1 - erf[(x-vt)/2(Dxt)^{\frac{1}{2}}]\} *$$
$$\{erf[(y+Y/2)/2(Dyx/v)^{\frac{1}{2}}] - erf[(y-Y/2)/2(Dyx/v)^{\frac{1}{2}}]\} *$$
$$\{erf[(z+Z/2)/2(Dzx/v)^{\frac{1}{2}}] - erf[(z-Z/2)/2(Dzx/v)^{\frac{1}{2}}]\}$$

where: $u = (v^2 + 4LDx)^{\frac{1}{2}}$

C - concentration in mass per unit volume of water

x,y,z - Cartesian coordinates where x coincides with the principal direction of ground water flow

t - time

Co - source concentration

erf - error function

v - average linear ground water velocity

Dx, Dy, Dz - principal values of the dispersion tensor

Y, Z - dimensions of the source in the y and z directions

L - first-order decay constant ($L = \ln 2.0 / t^{\frac{1}{2}}$)

$t^{\frac{1}{2}}$ - decay half-life

This equation describes the three-dimensional transport of mass from a continuous, finite (two-dimensional) source in a unidirectional flow regime.

To test the accuracy of the expanded equation, the ROKEY model and the numerical transport model DPCT (Schwartz and Crowe, 1980) were applied to a hypothetical contamination case. The results, shown in Figure 7, demonstrate a good agreement between the two. Domenico (1987) also has expanded the equation in a similar way to include first-order decay and has verified the accuracy of the revised equation.

The computer code developed for ROKEY was augmented so that changes in concentration at the source with time could be incorporated into a simulation. Also, provisions were made so that retardation due

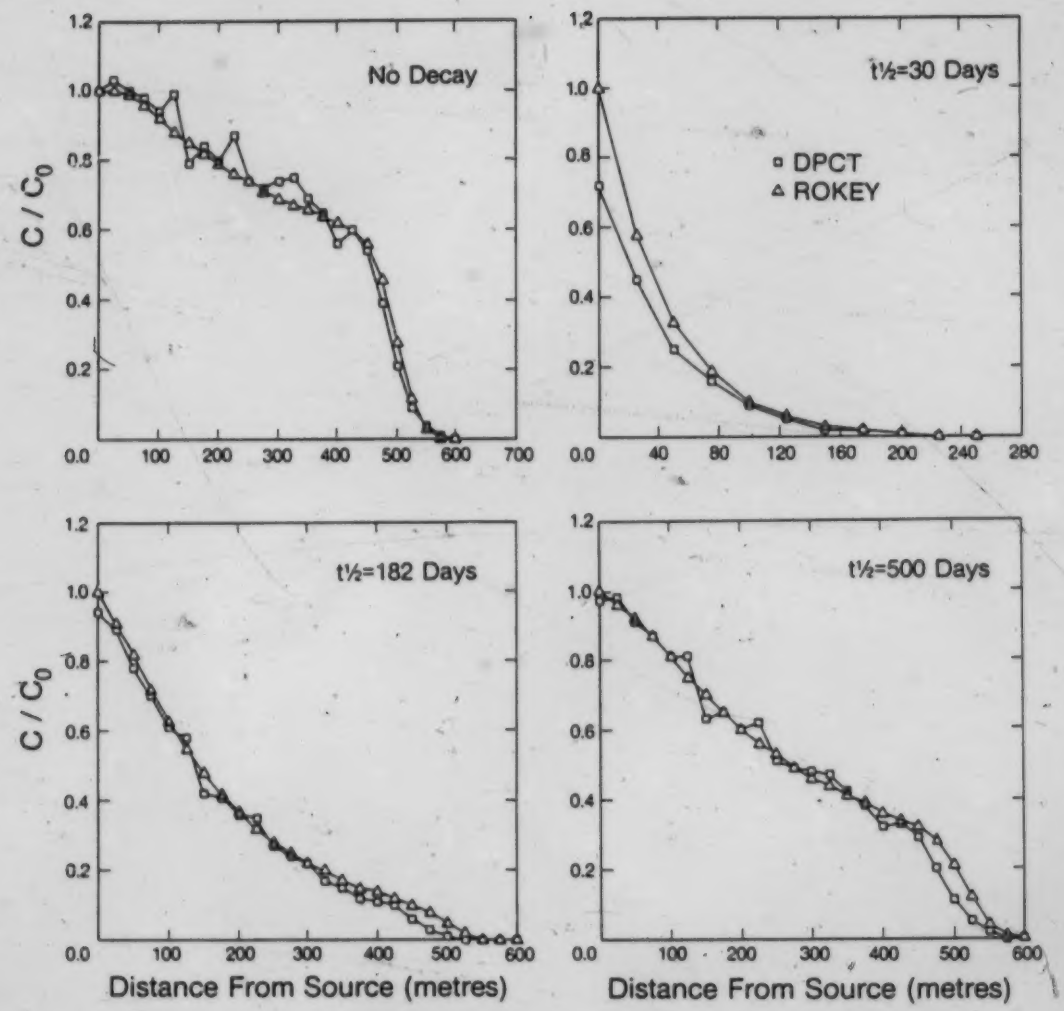


Figure 7 Comparison of ROKEY and DPCT Results

to sorption could be accounted for.

Subsequent testing revealed that the results provided by ROKEY would be most sensitive to the values assigned to the average linear ground water velocity (v), decay half-life ($t^{\frac{1}{2}}$), source concentration (Co) and the source dimension in the z direction (Z). Also, the value assigned to the organic carbon content of the transporting medium, which is used to calculate a retardation factor, would have a significant effect on the results.

4.3 The EXPAR Expert System

4.3.1 Description of the System

EXPAR was designed as a partitioned system with two levels of operation. At the top level is a set of nine computer forms that contain groups of related parameters and entry fields into which values are entered and displayed. Associated with each parameter is an elaboration program and an assistance program. The elaboration programs provide the parameter and tutorial information described in the previous chapter (Section 3.1.2.4). The assistance programs help the user in deriving values if assistance is requested. The complete set of elaboration and assistance programs represents the bottom level of the system. The use of the forms and access to the bottom level programs are controlled by the function keys located on the computer keyboard.

Deriving a set of input data with EXPAR and then running the ROKEY model is done in stages. At the start of a session, the user is given the opportunity to review the contents of the Introduction and Instructions modules. Entry into EXPAR comes next where the computer forms are displayed. The user enters the data for a site into the forms, seeking elaborations and assistance if and when the need arises.

Once completed, the pressing of a function key results in the storage of the data set into a file followed by the invocation of ROKEY. The user then can conduct another session with EXPAR, access the EXINS system or terminate the entire session. The results provided by ROKEY (a tabulation of concentration data) can be examined on the video screen or from a print-out. Optionally, a contoured plot of the concentration data can be obtained using a plotting package designed especially for ROKEY.

An option is provided for the user to work with an existing data set. The advantage of this feature is to avoid having to recomplete all of the forms when only one or two parameter values change in a transport simulation. Prior to the display of the forms, the user can identify a file containing the existing data set. The data in the file are entered into the appropriate fields on the forms, leaving the user free to make changes. This feature is useful when the distributions of several contaminants are modeled, or when the user wishes to conduct a sensitivity analysis with one of the parameters. Because a data set can be stored at any time, data from an interrupted session can be retrieved at a later time for completion.

4.3.2. The Top Level

The top level consists of three components: a control program, a global data base (GDB) and the set of computer forms (Fig. 8). The control program specifies the overall operation of a session, and controls the use of the GDB and the forms. The global data base stores the values entered for the (static) parameters displayed on the forms and any value derived for a parameter in an assistance program. Each piece of information accumulated during a session is represented as the

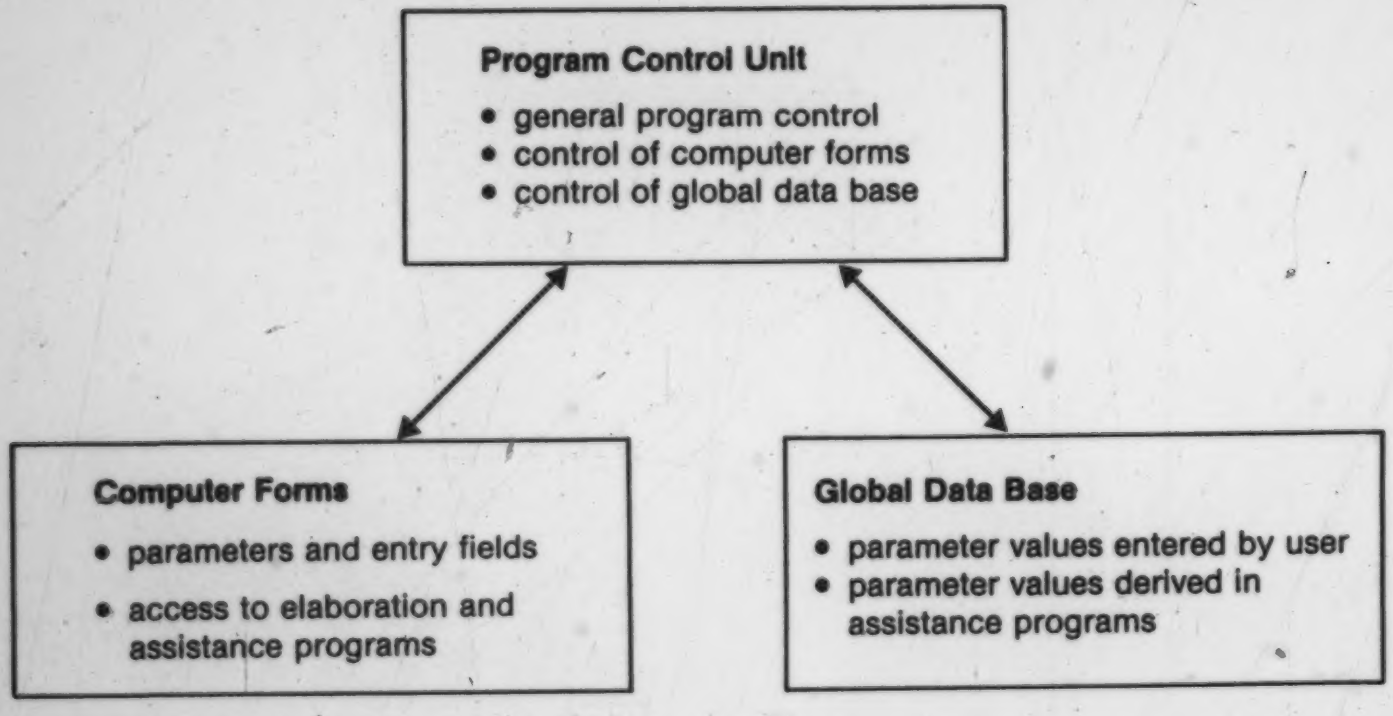

Program Control Unit
• general program control
• control of computer forms
• control of global data base
Computer Forms
• parameters and entry fields
• access to elaboration and assistance programs
Global Data Base
• parameter values entered by user
• parameter values derived in assistance programs

Figure 8 Top Level Architecture of EXPAR

value of a parameter. The contents of the GDB are transferred to a user-specified file before the exit from EXPAR.

The set of computer forms is given in Appendix B. Each form is divided into three regions (Fig. 9). The top region contains a message related to the use of the form or the parameters, such as whether values are required or optional. The middle region has a list of parameters and entry fields. The bottom region contains cues to the operations performed by the function keys. Briefly, F1 and F2 position the cursor within a form; F3 is pressed before the user makes an entry; F4, F7 and F8 access the various forms; F9 is pressed after all of the data have been entered (to run ROKEY); F10 terminates the session, returning the user to the computer's operating system. The F5 and F6 keys provide access to the elaboration and assistance programs.

Each form contains a group of related parameters, as indicated in Table 3. Forms 1 and 2 provide locations for the entry of general and specific bookkeeping information that is important in documenting the session. Forms 3 to 8 contain the flow, transport, source and chemical parameters. Two types of parameters are listed on Form 9: two subtasks and the output control data parameters. The details of the subtasks are described in Section 4.3.4.

There are three features of the top level that are particularly noteworthy. First, all entries made by the user for the parameters that require numeric values are checked for general validity using a set of 25 production rules. For example, EXPAR will accept values of effective porosity (expressed as a fraction) in the range of greater than 0 to less than 1.0. If the user enters a value that is outside of this range (i.e., a logically incorrect value), the video screen is cleared, a

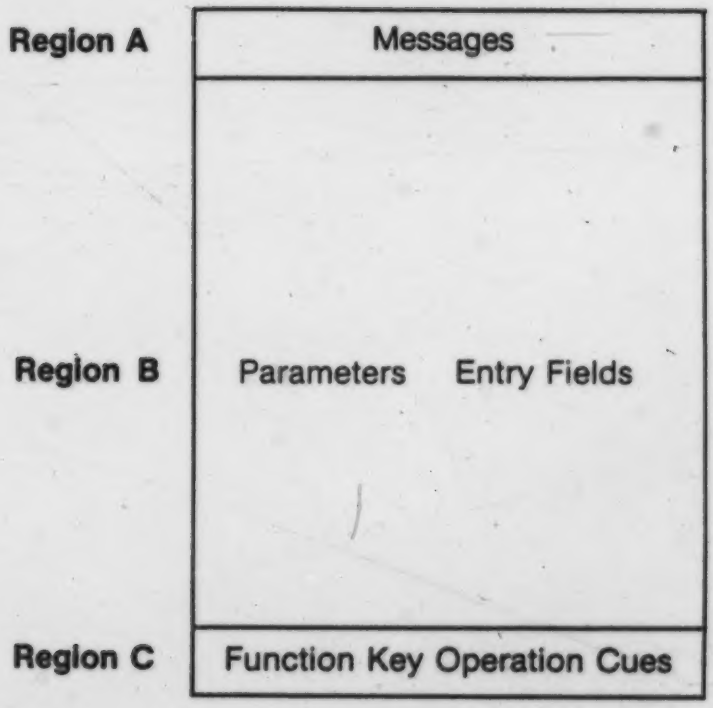


Figure 9 Computer Form Design

Table 3 Parameter Groups in EXPAR

Form No.	Type	Example
1	Bookkeeping parameters	Project No.
2	Bookkeeping parameters (and assignment of plausible and) default values)	Name of chemical
3	Basic transport parameters	Advection rate
4	Source size parameters	Depth to top of source
5,6	Concentration/time parameters	Initial concentration
7	Sorption parameters	Organic carbon content
8	Transformation parameters	Hydrolytic (abiotic) half-life
9	Output control parameters, subtasks	X interval, data set evaluation

message explaining the error is given and the form is redisplayed, but without the incorrect entry. Second, before storing the data set and invoking ROKEY, EXPAR uses a set of production rules to check that the minimum amount of data has been entered. Parameters for which values have not been assigned are reported after which Form 9 is redisplayed. This data-check insures that the operation of ROKEY will not fail, thereby eliminating the chance of a major disruption in a session. Third, with the exception of Form 2, a session with EXPAR can be conducted at the top level. The bottom-level programs are available on an "as needed" basis only. The uniqueness of Form 2 is discussed in Section 4.3.4.

4.3.3 The Bottom Level - Elaboration Programs

Most elaboration programs consist of two components: a control block and one or more text files (Fig. 10a). However, several programs make use of information that has been placed in the GDB.

Information that is specific to a parameter is given in one text file. Several files of tutorial information can be accessed through the programs for six parameters. For example, tutorial files for mechanical dispersion are included in the program for the parameter "Dispersivity - longitudinal (x)". Dispersion is a complex topic and many users will not have a complete understanding of the process. In order to present the tutorial in an effective manner, 13 text files are available. All of the files are accessed through a menu. Because it is expected that users with varying levels of knowledge will be using the system, the files have been subdivided into introductory and advanced levels. In addition to mechanical dispersion, tutorials are given for advection, diffusion, sorption, abiotic hydrolysis and biotransformation.

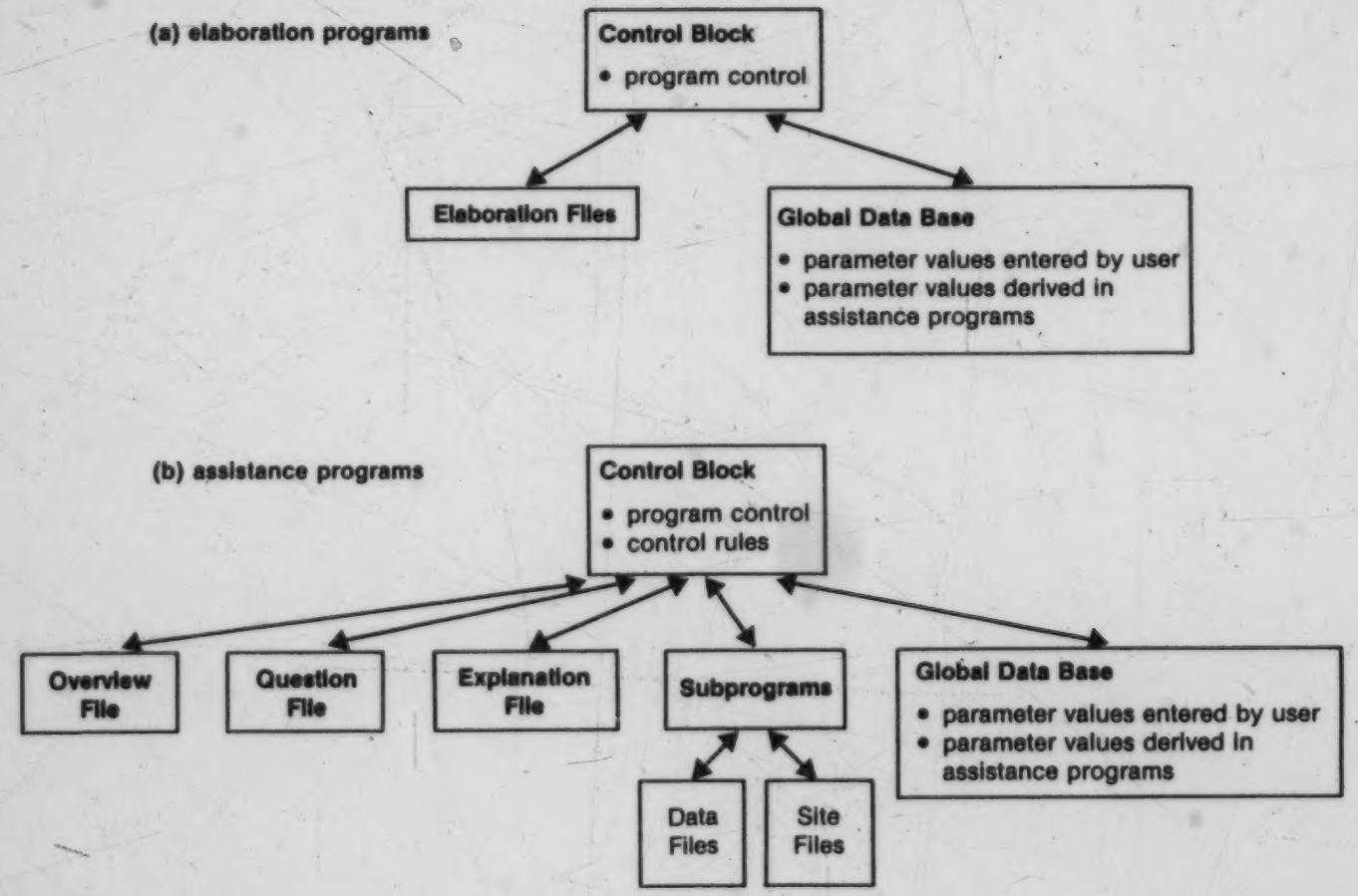


Figure 10 Bottom Level Architecture of EXPAR

The elaboration programs are designed to serve two purposes. The first is to help the user assign parameter values. The second is to educate the user by providing in-depth, practical knowledge about a parameter and, where appropriate, about the transport process involved. Consequently, users who have become familiar with the files should not only be proficient at using the ROKEY model but, also, they should have a better understanding of contaminant transport. Lists of the topics discussed in the tutorials are presented in Appendix C.

4.3.4 The Bottom Level - Assistance Programs

The assistance programs are, in essence, individual expert systems. These programs operate by attempting to derive a parameter value (e.g., Dx, foc) starting with the most-preferred method and ending with the least-preferred. In most cases, this latter method involves the assignment of a default value. A diagram of the general architecture of the programs is shown in Figure 10b. It should be noted that due to the uniqueness of each parameter, varying types and amounts of assistance are available. Consequently, not all of the components shown in Figure 10b are required for each program. A program is carefully tailored to the parameter and, as such, only the components logically required to give assistance are included.

The operations that take place in an assistance program are coordinated by the control block. Upon entering a program, the control block first processes any control rules that may be present. These rules can perform two functions: (1) determine whether supporting data that may be required for offering assistance have been entered into the GDB, and (2) determine which part of the control block is to be accessed next. In the case of (1), the user will be returned to the

forms so that data which support the program can be assigned. For example, providing assistance for dispersivity values is dependent on the scale of the transport problem. This can be determined when the advection rate and the simulation period are known. The user is required to enter values for these two parameters before assistance with the dispersivity values can be provided. For (2) above, the assistance provided may be dependent on one or more factors. As an example, the assistance for longitudinal dispersivity will depend on whether the scale of the transport problem is local, global or regional. If it is regional, questioning (described below) related to this scale will be accessed, thereby bypassing the questions related to the other two scales.

Once the control rules, if any, have been processed the control block accesses a file that gives a brief overview of the factors involved in deriving a value. Also discussed is the reasoning that will be employed by the program. After this, a QA (question-answering) network is accessed in which a series of questions related to the parameter is presented. The user can respond with either "Y", "N", "?", "F", "B" or "E" which represent yes, no, explanation, forward, back and exit, respectively. A "Y" or "N" entry will lead the user to another question or to a subprogram. A "?" response is similar to an F5 entry made at the top level - that is, the video screen will be cleared and an explanation will be given as to (1) the meaning of the question, (2) why it was asked and (3) the factors related to the information being sought. Also, if the question relates to a specific hydrogeological test or method of analysis, a brief description of this will be given plus a list of references to the literature. When the user is finished

viewing the explanation, the screen is cleared and the originating question is redisplayed. An entry of "F" or "B" allows the user tc move forward or backward rapidly through the questions. This feature is intended for users experienced with the program who wish to access a specific subprogram without having to respond to other questions. An entry of "E" will result in the exit from the program and the return to the originating form. The QA network is a simple decision tree. Each question is a node in the tree with the yes and no responses leading to other questions or to subprograms.

There may be up to four types of subprograms available in an assistance program. These are designed to:

(a) read a value entered by the user;

(b) calculate a value using a conventional computer program;

(c) display tables of values and summaries of case studies;

(d) recommend a default value.

The reading of a user-volunteered value in (a) is essentially the same as accepting a value on a computer form. In the latter case, the user may have only one value for a parameter and will likely enter it on a form. However, if more than one value is available and the user is unsure of which one is appropriate, the subprogram will make a recommendation and then read the user's chosen value.

Several subprograms are conventional computer programs. Some of these were prepared especially for use in EXPAR while the remainder were developed previously by others and adapted for use in the system. All of the programs calculate the value of some parameter. The programs range from very simple (e.g., solving Darcy's equation), to moderately advanced (e.g., solving an equation that calculates a hydraulic

gradient value using a linear interpolation scheme) to very advanced (e.g., the GENPAR1 code [Milne-Home, 1985] that calculates a value for hydraulic conductivity using a variation of the Kozeny-Carman equation). Any value derived by a conventional program is displayed to the user who can either accept or reject it. The assistance program provides help with this decision. However, the final decision is left to the user's discretion.

The assistance programs for five of the parameters contain data files that list values measured or derived at specific sites. The user, upon examining these files, may choose to use one of the values if the site is similar to the one presently under investigation. To help in assessing the similarity of the sites, the user can access a site file that contains a summary of the information for the site. If the user is satisfied that a value in the data file is appropriate, he will be asked to enter it. The data files have been set-up so that the user can incorporate information from his past projects, thereby creating customized files. In addition, any number of site files can be built. This arrangement was developed as a convenience for the user by providing a location within the system for the storage and retrieval of his own data.

The fourth type of subprogram recommends a default value. The default value may have been made known to the system from an entry made on Form 2 (discussed below) or it may be coded in the subprogram. Some of the values are based on public knowledge, such as hydraulic conductivity and porosity values given in textbooks or manuals. However, some are based on my heuristic knowledge. Examples include (1) the concentration of the chemical at the source, (2) organic carbon

contents of various types of porous media, (3) a regional scale value for longitudinal dispersivity, and (4) the thickness of the mixing zone at the source. Developing the heuristic values during the project was unavoidable because typical values for some of the parameters do not exist in the literature. The user is required to accept the default value or enter a value within a specified range.

All parameter values entered by the user within a subprogram or generated by a conventional program can be checked for consistency with the data in the GDB before leaving the program. Any inconsistency detected is reported along with one or more recommended actions. These actions include the cancellation of the value and either continuing with the questioning sequence or reimplementing a conventional program. These alternatives may be initiated automatically by the program, if the user agrees, or conducted manually by the user by selecting a new position in the questioning sequence. If an inconsistency is not detected, the value will be entered into the GDB along with any values derived for parameters that are specific to the subprogram. Finally, the user is returned to the originating computer form with the value placed in the appropriate entry field. The user has the freedom to change the value at any time by either changing the value directly on the form or by reentering the assistance program.

The assistance programs for the parameters listed on Form 2 are different than the others. In the case of Form 2 itself, users can not make entries directly but, rather, they must use the assistance program. The assistance programs for 'Name of chemical', 'Type of source', 'Type of stratum material' and 'Slope of ground surface' contain menus from which the user makes the appropriate selection. For

example, the assistance program for 'Name of chemical' will display a list of the organic chemicals tabulated on the U.S. EPA's 129 Priority Pollutant list. Associated with each chemical is an index number. If the chemical to be modeled is on the list, the user will enter its index number and both the name of the chemical and the index number will be entered on Form 2. This procedure also results in the placement of plausible values for the chemical's solubility and octanol/water partition coefficient into the GDB if available in one of the system's data bases. Additionally, available default values for hydrolytic and biotransformation half-lives are entered into the GBD. Similar support facilities are given for the other three parameters noted above, and include various text files, data bases and production rules designed to provide default values for the parameters listed on the other forms. Plausible and default values are entered into the GDB automatically. Plausible values will be entered into the entry fields on the other forms when they are displayed. Default values may be used later in the session for consistency checking and for providing assistance. The menu choices available for the Form 2 parameters are reasonably comprehensive but the user can make his own entry if an item is not in a menu.

The two parameters listed at the top of Form 9 are subtasks. The first subtask performs a check on the consistency of the data entered on the previous forms and is accessed when the user presses F6. Any inconsistency and suggested corrective actions will be reported. Normally, a user will use this feature to make sure that errors were not made. For example, an advective velocity of 0.01 cm/s assigned to a dense clay will be flagged and reported. The user can ignore any

suggested corrective action. This is an important characteristic of EXPAR. The user is in control at all times and can ignore any advice or decision given at any point in a session. The second subtask, also enacted when F6 is pressed, calculates the approximate length of the contaminant plume. This feature may be useful during sensitivity analyses when the user would like to quickly assess the impact of changing the value assigned to a parameter without having to run the ROKEY model. Also, it serves as an indirect means of assessing the overall accuracy of the data set.

A summary of the contents of the assistance programs is given in Appendix D.

4.3.5 Design Considerations

At the start of the project several objectives related to expert systems design and practical software considerations were defined for EXPAR. Perhaps the most important was to develop a fast and effective computer program that would be of use not only to non-experts, but also to experts and all users who have become familiar with EXPAR. Meeting this objective required the design of interfaces that would allow the rapid entry of data and quick access to all parts of the system. It was with this in mind that EXPAR was built as a partitioned system, relying mainly on computer forms, short text files and QA networks to communicate with the user.

For non-experts, EXPAR assists in preparing a reasonable set of input data for a given site. The extensive elaborations should increase the user's understanding of the model parameters and contaminant transport in general. For the expert and those users who have become well-acquainted with the contents of EXPAR, the system can be used like

standard modeling software. Because the advice, assistance and elaboration components are kept in the bottom level, these users can fill in the top level forms quickly and run the ROKEY model without any disruptions or interference. The bottom level programs are available when needed. The only exception is the assistance programs for the Form 2 parameters.

The choice of computer forms as the main method of data acquisition was an obvious one. Using forms avoids having to answer over fifty questions as would be the case if all of the interaction was conducted with a QA scheme. From a human engineering perspective, forms are a natural way to collect data. The most convenient interface would have been one form such as the video display and the master panel used in ONCOCIN and ELAS. However, this approach would have required the use of a large graphics terminal rather than the video screen that comes with a microcomputer.

Using QA networks in the assistance programs, however, was the optimum method of extracting information from the user. This approach easily handles the dynamic variability of the information. For example, a question presented at any point may depend on the user's earlier responses in the program or on a form. Consequently, a structure like a QA network provides the necessary control whereas static structures such as forms and menus do not.

One of the objectives in designing the QA networks was to allow fast and easy access to the conventional programs embedded in some of the assistance programs, either during a regular session with EXPAR or whenever the user had a need to use a conventional program. As an example of the latter, the user may wish to process grain-size data

with the Kozeny-Carman program while working on a project unrelated to contaminant transport. The appropriate subprogram can be accessed quickly with the forward ("F") and back ("B") responses to questions in the QA sequence for the appropriate assistance program. Once the user has finished working with the code, an entry of "E" to the next question in the sequence will result in the return to the computer form.

Two expert systems features that were intentionally avoided during the course of the project were natural language understanding and the use of certainty factors. In systems with natural language understanding, users are able to communicate with the program using a restricted version of English. Although there is a great appeal to an interface like this, the time and effort involved in developing it are prohibitive. Also, systems that use restricted versions can be slow and tedious because the user must remember which words or phrases are acceptable. Although not as appealing, the entry of data in the computer forms and the simple responses in the QA networks are effective means of communication.

Certainty factors can be used to express the degree of confidence in a result or conclusion reached by an expert system. However, including these in a system is restricted to those problem domains in which it is feasible to assess the accuracy of some or all of the data and knowledge that will be incorporated into the system. Assigning meaningful certainty factors to hydrogeological data and the results given by transport models is difficult. This problem precluded the assignment of degrees of confidence to the input data prepared in EXPAR and the corresponding results provided by ROKEY. One of the most

distinguishing characteristics of hydrogeology as a science is the large number and size of the uncertainties associated with almost all of the phenomena involved. The sizes of the potential errors associated with the parameter values required by ROKEY range from tens of per cent (e.g., porosity) to orders of magnitude (e.g., advection rate).

Although not a specific design consideration, the choice of computer language can have a major bearing on the construction of a system. Most expert systems are programmed in either LISP or PROLOG. However, EXPAR was written in FORTRAN for three main reasons. First, it is advantageous to write any computer program in one language. The conventional programs included with several of the assistance programs and the ROKEY model were written in FORTRAN. Therefore, it was easier to program EXPAR in FORTRAN rather than to write the entire system in another language. The advantage of selecting FORTRAN extends further. Expert systems are modified frequently and it is possible that conventional FORTRAN codes could be added or substituted at a later time. When this occurs, inter-program communication will not present any difficulties. The second reason relates to programming familiarity. It is anticipated that the programming methods and structures developed for EXPAR could serve as integral components of future expert systems developed by other hydrogeologists. Most hydrogeologists work only with FORTRAN and, as such, very few would be interested in learning another language. Third, LISP and PROLOG did not offer any advantages over other languages. All of the knowledge representation structures and procedural components could be written in virtually any language.

4.3.6 Discussion

EXPAR is a multi-level expert system in two respects. First, it

serves as a surface level model for ROKEY, which according to Hart's analogy is considered to be a deep level model of reasoning. Communication between the two is achieved through the global data base but, unlike ELAS, the transfer of information is in one direction rather than in two. Second, like ELAS, EXPAR includes existing software that also are deep level models of reasoning. Several of the components in EXPAR act as surface level models of reasoning that provide advice as to when the conventional programs should be used. Assistance also is given to guide their proper use and to filter the results that they provide. The ability to integrate existing software into EXPAR demonstrates the flexibility of the programming constructs. More importantly, embedding these programs into an expert system increases the power and utility of both the system and the programs.

One of the fundamental objectives in designing an expert system is to make it sufficiently flexible so that modifications can easily be made during the stages of development and afterwards as the system is maintained. Flexibility is achieved by employing a simple architecture and simple, modular knowledge representation structures. In a rule-based system, the architecture consists of only three components (program control unit, knowledge base and global data base) and one simple type of knowledge structure (production rules). Partitioned systems also are based on simplicity and modularity but the advent of a more practical interface (i.e., computer forms) has given rise to additional knowledge representation structures. In the ELAS system, a master panel (computer form) that lists the static parameters and menus (a type of control block) are used in conjunction with production rules. In ONCOCIN, associated with the video display (computer form)

are data blocks and parameters, while the control and reasoning operations are performed with control blocks and production rules.

Simplicity and modularity in the basic architecture of EXPAR were achieved through the use of computer forms, a global data base, and the set of elaboration and assistance programs. Each computer form contains a logical grouping of parameters, and the entry of user-volunteered information is governed by a set of production rules. The program control unit was designed so that forms could be added or deleted easily. Also, modifications to the contents of the forms can be done with little effort. The set of data in the GDB represents a data block which can be stored and retrieved for additional processing. The elaboration and assistance programs are modular units which, if the need arises, could be replaced by different programs without difficulty. Communication between these programs and the computer forms is achieved simply by the connection of all of these components to the GDB.

Within the assistance programs, control blocks and QA networks serve as convenient structures for organizing the knowledge into hierarchies that represent reasoning strategies. The knowledge itself consists of simple units in the form of question-explanation couples, rules, text files and conventional programs. Any of these can be revised with relative ease. The incorporation of all of the above structures into EXPAR has resulted in a flexible and expressive system.

4.4 The EXINS Expert System

4.4.1 Description of the System

The EXINS system is divided into two main sections (Fig. 11). The first section is composed of three tutorials that describe several

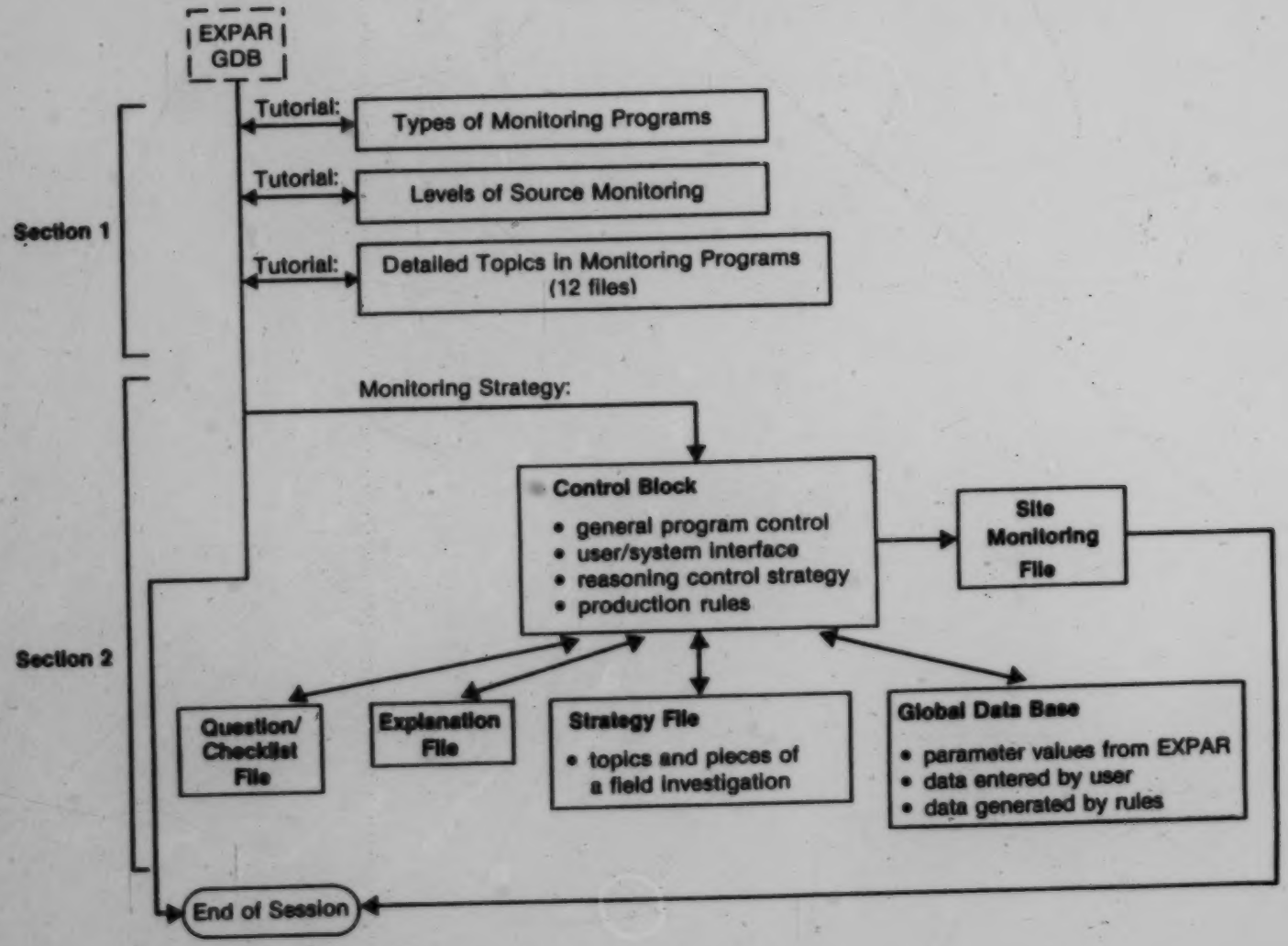


Figure 11 Architecture of EXINS

aspects of monitoring programs. The first one gives a short summary of the four types of monitoring programs (i.e., ambient, source, case preparation and research). The second one presents a description of the detection and interpretive levels of source monitoring. The third tutorial contains summaries of the 12 topics that are usually considered in the first stage of a monitoring strategy (Table 2). In addition to tutoring the user, the information in the files serves to provide justifications and explanations for the particular strategy to be recommended in the second section. Associated with the topic "Analytical Program" are files that list the chemical parameters and constituents commonly selected for analysis for nine types of sources. These source types are listed in Table 4. All three tutorials are accessed through the user's responses to questions while access to the individual topics and categories indicated in Tables 2 and 4 is menu-oriented. After viewing the tutorials, the user has the option of ending the session or continuing with Section 2.

The preparation of a monitoring strategy for the site previously modeled with ROKEY is undertaken with the six components that represent Section 2 of EXINS. Five of the components (Control Block; Question, Explanation and Strategy Files; Global Data Base) interact with the user to prepare the strategy which is written to the sixth component (Site Monitoring File) once it has been formulated. The most unique feature in EXINS is the knowledge representation structure referred to as the Strategy File. This component is a data file that contains an indexed listing of over 70 "pieces" that can be used to build a strategy. All of the pieces are associated with the 12 topics noted in Table 2 and are grouped according to the topic to which they apply. For

Table 4 Types of Sources Considered in EXINS

Category No.	Type of Source
1	Municipal Landfills
2	Municipal Sewage Lagoons and Ponds
3	Domestic Septic Tank Systems and Cesspools
4	Industrial Wastewater Lagoons and Ponds
5	Industrial Septic Tank Systems and Cesspools
6	Industrial Landfills - all wastes are known
7	Industrial Landfills - all wastes are not known
8	Sources of Leaks and Spills - crude oil and derivatives
9	Sources of Leaks and Spills - other chemicals and materials

example, the topic "Locations of Monitoring-Well Nests" contains three pieces (Fig. 12) which are used by the system to recommend the number and configurations of nests for minimum detection, detailed detection and interpretive programs. The complete Strategy File is given in Appendix E.

A strategy is built through the selection of the one piece in the group of pieces for each topic that is best-suited for the site. This is accomplished with a set of production rules. Once a rule has selected a piece, the corresponding index number is stored in the global data base. At the end of the session the index numbers for 12 pieces will be present in the global data base. The final operation in the session involves the writing of the strategy to the Site Monitoring File. This is accomplished in the Control Block where the index numbers in the global data base are compared with those in the Strategy File. For numbers that match, the corresponding pieces are copied and written to the Site Monitoring File. The user then can examine the file by having it displayed on the video screen or by obtaining a printed copy.

The interaction between the user and the system involves the display of a series of questions and checklists to which the user makes responses that are appropriate for the site under investigation. Each response first is entered into the global data base and then processed by the rule set. A successful match between the response and a rule leads either to the display of another question (or checklist), or to the entry of an index number into the global data base. For each question and checklist presented, the user has the option of entering a "?" response. This will result in the display of an explanation that outlines the reasoning being employed, followed by the redisplay of the

```
0062) Locations of Monitoring-Well Nests
006   ---------------------------------
006
000 The info. for 007 is for a minimum detection program
000
007   The suggested configuration is:
007
007        Number of A nests - 1
007        Number of B nests - 1
007        Number of C nests - 3
007
000 The info. for 008 is for a "detailed" detection program
000
008   The suggested configuration is:
008
008        Number of A nests - 4
008        Number of B nests - 3
008        Number of C nests - 3
008
000 The info. for 009 is for an interpretive program
000
009   The suggested configuration is:
009
009        Number of A nests - 4
009        Number of B nests - 3
009        Number of C nests - 10
009
```

Figure 12 Pieces in "Locations of Monitoring-Well Nests" Topic

question or checklist. As indicated in Figure 11, the global data base in EXPAR is transferred to the EXINS system. Information from this is used to help determine whether a detection or an interpretive program is most appropriate for the given site.

Examples of two rules are given in Figure 13. All of the rules were written as If-Then-Else statements in FORTRAN 77 which, if converted to the If-Then syntax of standard production rules, would number in the order of 130.

4.4.2 Discussion

The architecture of Section 2 in EXINS (Fig. 11) is a modified version of the architecture for a rule-based system (Fig. 2). The main differences between the two are that with EXINS the production rules are located in the Control Block rather than in the knowledge base, the knowledge base has been separated further into a Strategy File and files for questions and explanations, and the various "conclusions" that can be offered by the system are expressed as groups of pieces in the Strategy File.

Placing the rules in the Control Block as programming statements enhances the organization of the system. In a traditional rule-based system, the rules can not be organized in a way that would identify groupings of similar rules. There may be implicit groupings of rules that apply in specific situations and at certain stages in a session, but there is no explicit indexing of the rules by situations or stages (Aikins, 1983). There are two main disadvantages to this latter method of organizing rules. First, it usually is not possible to determine the potential reasoning sequences that can occur by examining the knowledge base. Second, it is usually difficult to make modifications or

If: the source is an industrial wastewater lagoon

Then: ask the user whether all of the chemicals have been identified

If: a detailed detection program strategy is required

Then: enter index no. 008 into the global data base

Figure 13 Two Rules in EXINS

additions to sets of rules without explicit groupings, particularly where changes may have indirect effects on other rules. These problems were minimized in EXINS by organizing the rules into explicit groups according to the strategy topic to which they applied. As a result, the system was built on a topic by topic basis with the programming for each appearing as modular units. Instead of working with a group of about 130 rules, only the rules associated with a topic (ranging between 2 and 36) had to be considered at any one time. The reasoning associated with the selection of the strategy piece for each topic can be followed easily and modifications to the rule sequences can be made with little trouble. Any future changes to the rules would require the recompilation of the code but it is anticipated that changes more likely would occur in the Strategy File.

The strategy pieces contained in the Strategy File represent the consequents of rules and are linked to the rules with index numbers. The Strategy File was created to provide a convenient means of expressing the consequents of the rules. Without it, the strategy pieces would have to be written into the rules, thereby adding more programming to the Control Block and making revisions to the pieces more difficult. The Strategy File exists as a text file that is external to the programming and, as such, revisions can be made readily by the user with a standard editor if the need arises. The file is fully commented to assist users with this.

5. SYSTEM EVALUATION

5.1 Introduction

Some type of empirical approach is usually used to evaluate the performance of an expert system. A common one involves the comparison of results generated using the system with results obtained using an independent approach for the same problem (Weiss and Kulikowski, 1985). In one testing procedure, one or more nonexperts could apply the system to a problem with the results compared to those obtained by an expert. A variation of this approach would involve replacing the expert by a correct answer that is available in the form of a measurable, physical entity.

An evaluation of Expert ROKEY (prior to the development of EXINS) was conducted using this latter approach. Eighteen students in a graduate-level course in hydrogeology were given a description of an existing contamination problem and were assigned the task of utilizing the system to predict the distributions of three organic chemicals. The evaluation that followed was centered on assessing the similarities between the simulated and known plume configurations. In addition, answers to the following were sought:

(a) was the heuristic knowledge in EXPAR valid ?

(b) where the results obtained by the subjects sufficiently accurate for risk assessment decisions or for planning the first stage of a monitoring strategy ?

(c) was the system sufficiently user-friendly so that it could be used without difficulty by all of the subjects, regardless of computer ability ?

(d) to which parameters were poor or incorrect values given ?

As with all evaluations, my main goals were to assess the present capabilities of the system and to identify those parts that required improvement.

5.2 Site Description

The site selected for the evaluation was the Special Waste Compound, part of the Gloucester Landfill, located seven kilometers south of Ottawa, Ontario (Fig. 14). A detailed description of this contamination problem is presented in Jackson et.al. (1985) and additional information is available in a consultant's report (Graham et.al., 1985). The following was summarized from these two sources.

The Special Waste Compound occupies an area of approximately 0.36 hectares (0.9 acres) in the southwest corner of the municipal landfill. Over a 12-year period (1969 - 1980), a variety of inorganic and organic wastes were disposed of in shallow trenches. In some instances, the wastes were ignited before burial. Available records show that at least 108,000 litres of liquids, and 24 tonnes (24000 kg.) of boxed liquids and solids were buried. As is the case for many disposal sites, a complete inventory of wastes and the dates on which disposal occurred are not available.

The site is underlain by glaciofluvial and littoral deposits of Quaternary age. The uppermost bedrock unit is an Ordovician limestone. Five hydrostratigraphic units have been identified on the basis of hydraulic properties (Fig. 15). Although the disposal trenches are in contact with Unit E, the majority of contaminant migration is in Unit C. Downward leakage into this deeper unit is caused by discontinuities in Unit D which act as "hydraulic windows". Once in this lower aquifer,

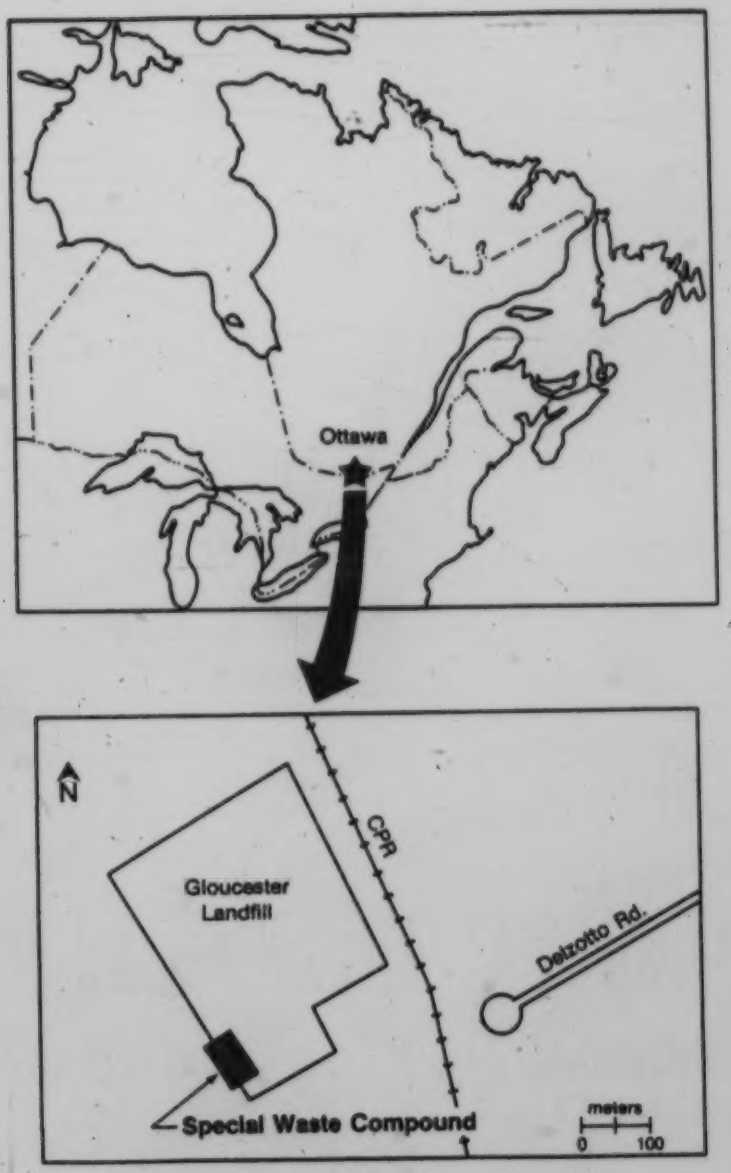


Figure 14 Location of the Gloucester Landfill

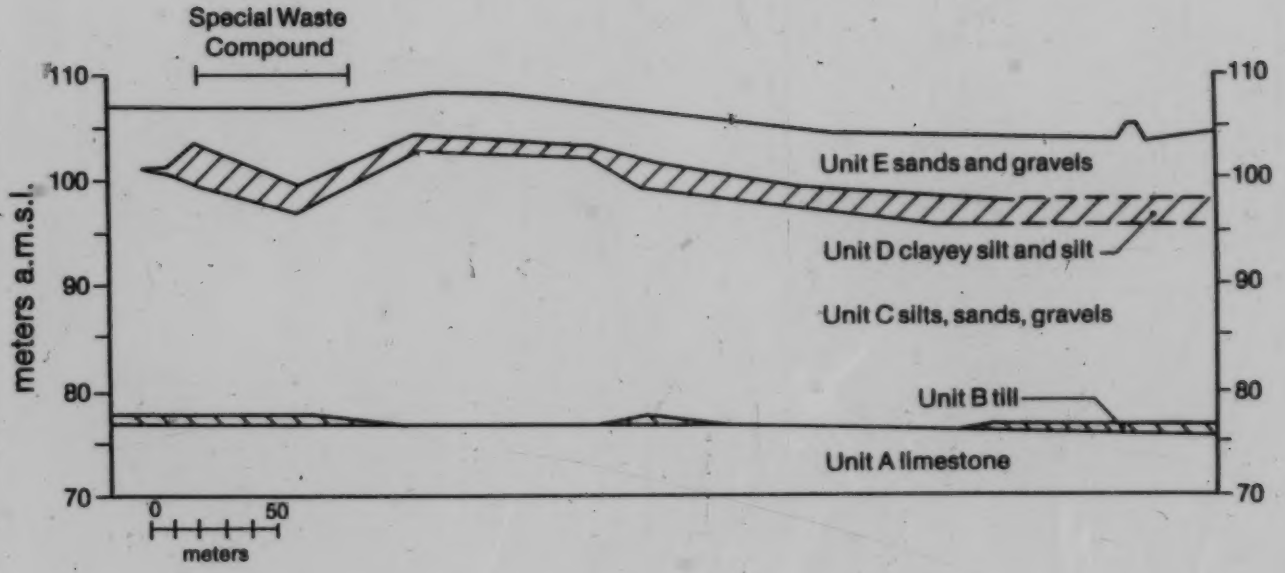


Figure 15 Hydrostratigraphic Units at the Gloucester Landfill
(adapted from Jackson et.al., 1985)

migration is predominantly horizontal in an easterly direction. Unit C is a heterogeneous stratum. It is comprised of a complex sequence of interstratified and interfingered layers of silt, sand and poorly-sorted gravel. As expected, the hydraulic characteristics are quite variable. Hydraulic conductivity values, as determined from the results of slug tests, range over two orders of magnitude (avg. 0.011 cm/s). Aquifer tests yield a smaller range with an average value that is higher (0.049 cm/s). The results of point dilution tests and an analysis of the distribution of bomb tritium yield a ground water velocity of about 5 cm/day. The most important mineralogical parameter associated with the calculation of relative velocities (i.e., retardation factors) for organic chemicals is the organic carbon content. For Unit C, values for this parameter in the range of 0.0015 to 0.0131 (expressed as fractions) have been determined.

Three of the organic compounds disposed of at the site and selected for the evaluation exercise are 1,2-Dichloroethane [DCE], trichloroethylene [TCE] and acetone [ACE]. The plumes for these compounds and for Cl^- (based on 1983 data) are shown in Figure 16a. The maximum migration distances of DCE, TCE and ACE are estimated to be 60, 90 and 460 meters, respectively. Acetone appears to be a non-retarded compound in that its distribution is the same as that for Cl^-. Based on the Cl^- data, it appears that both Cl^- and acetone were introduced into the subsurface in 1969 when disposal operations first began. The plumes for DCE and TCE are much smaller due to one or a combination of three factors: sorption, transformation reactions, and disposal at a time later than 1969.

It can be expected that estimating the distributions of these

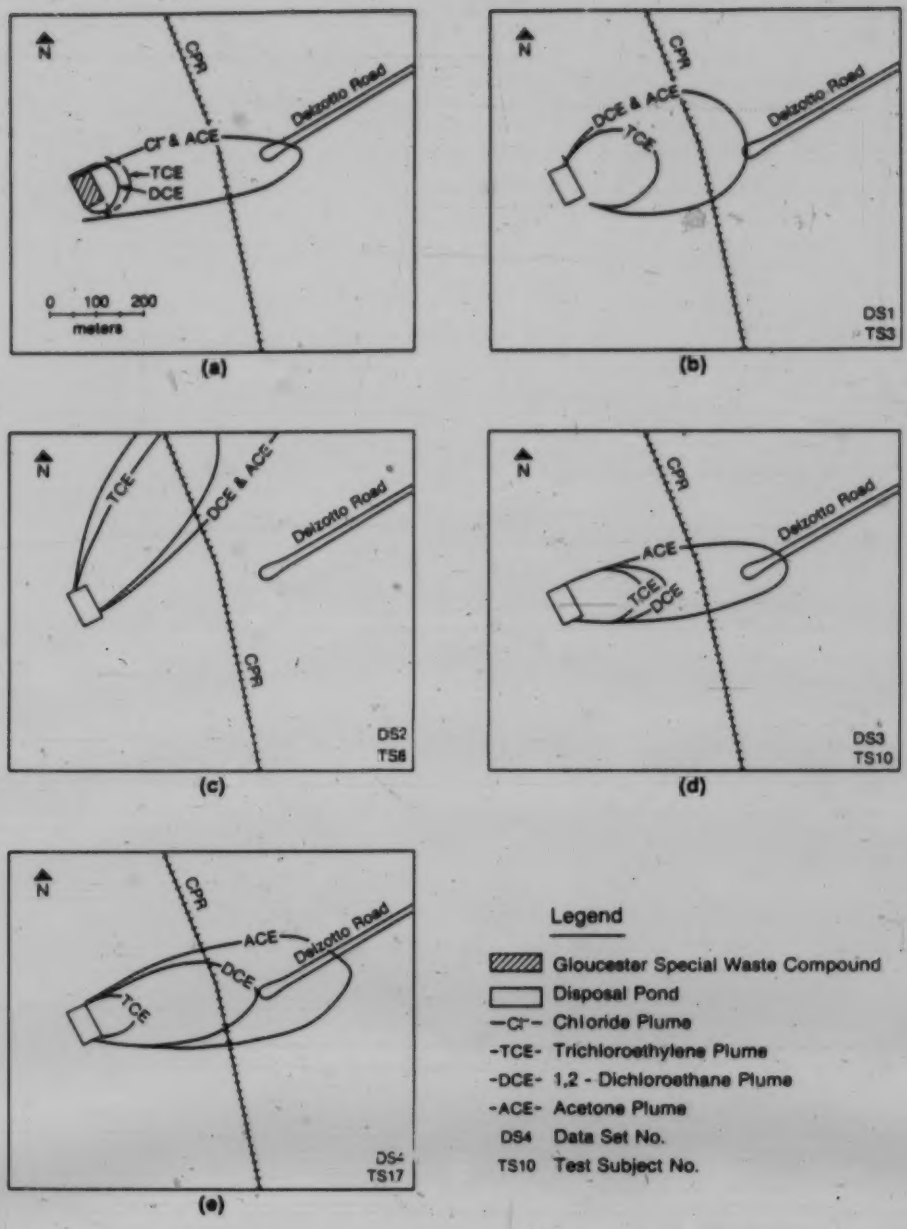


Figure 16 Actual and Simulated Plume Configurations

compounds with any transport model could be a difficult task in view of the complex nature of the problem. In the case of the data requirements of ROKEY, the absence of historical information for the disposal operation precludes the assignment of actual initial concentration values. It also rules out the possibility of assigning changes in source concentration with time which is certain to have occurred. The assumption inherent with ROKEY that transport occurs horizontally from the source in one homogeneous stratum is not met. As a result, applying Expert ROKEY to the Compound setting presents a significant challenge to both EXPAR and ROKEY. Specifically, a reliance on heuristic knowledge would be unavoidable and several assumptions inherent with the transport model would not be closely met.

5.3 Description of the Exercise

The 18 test subjects were informed that an unidentified facility had been discharging a variety of organic wastes into a small unlined pond for 15 years. They were given a site map showing the location and surface area of the pond, the ground surface topography of the area, and the lithologs for two boreholes. Each subject also was given one of four data sets that contained site-specific information (Table 5). All were assigned the task of using Expert ROKEY to predict the distributions of DCE, TCE and ACE. Once completed, they were to provide detailed accounts of the methods and reasoning used in deriving the values for the parameters in EXPAR.

The information given to the subjects was taken from the report by Jackson et.al. (1985). The data sets were prepared using the information for Unit C. To simplify the source characteristics, the Special Waste Compound was described as being an unlined pond rather

Table 5 Data Sets for Evaluation Exercise

Data Set No.	Data
1	- grain-size distribution data for one sample
2	- one hydraulic conductivity value based on results of an aquifer test - hydraulic head and concentration data for three piezometers
3	- grain-size distribution data for one sample - one hydraulic conductivity value based on results of an aquifer test - hydraulic head, hydraulic conductivity and concentration data for seven piezometers (hydraulic conductivity values determined from results of single well response tests)
4	- two hydraulic conductivity values determined from results of single well response tests - the mean ground water velocity value based on results of point dilution tests - hydraulic head data for six piezometers - one organic carbon content value

than a series of disposal trenches. The subjects had to assume that the organic chemicals were entering the subsurface on a continuous basis over the 15-year period from 1969 to 1983. This assumption likely describes the source loading of acetone reasonably well. However, there is no confirmation that the loading of DCE and TCE was continuous for 15 years.

5.4 Results and Discussion

The plume configurations simulated by four of the test subjects, each working with a different set of site-specific data, are shown in Figures 16b, 16c, 16d and 16e. I have examined only four sets of results due to the large number of plumes and associated parameter values generated in the exercise. The subject selected from each data-set group was chosen on the basis of the individual who met the minimum qualifications of users (discussed in Chapter 3) the closest. Test Subjects 3 and 17 (Figs. 16b and 16e) are practising hydrogeologists experienced in ground water resource development but have little experience with contamination problems. Test Subjects 8 and 10 (Figs. 16c and 16d), a practising geologist and a graduate student in hydrogeology, respectively, have educational backgrounds in hydrogeology but no relevant experience.

A comparison of the simulated acetone plumes with the actual plume configuration (Fig. 16a) illustrates that the plume in Figure 16d is a close match, those in Figures 16b and 16e can be considered to be reasonable estimates and the plume in Figure 16c bears no resemblance. A review of the parameter values showed that the lengths of the simulated plumes differ mainly due to variations in the values assigned to the advection rate and to the source concentration.

The acetone plume shown in Figure 16c is a poor approximation of the actual plume both in terms of size and orientation. The large plume size was an artifact of the particular data provided. The average hydraulic conductivity value given in Data Set No. 2 (DS 2) was derived from the results of the aquifer tests. However, this value is an overestimate of the actual value. The use of this larger value in conjunction with the values derived by the subject for hydraulic gradient and effective porosity (both of which were appropriate for the site) lead to an advection rate of 11 cm/day. This value is more than twice the actual value of 5 cm/day noted previously. The error in the transport direction was caused by the subject's misuse of the given hydraulic head data. EXPAR does not provide assistance in deriving the principal direction of flow (i.e., the transport direction) from hydraulic head data. Therefore, the user is responsible for its derivation. Test Subject No. 8 determined the direction indicated in Figure 16c using a three-point solution. However, this should not have been attempted because the three points virtually form a line.

The selection of acetone forced the subjects to seek advice from EXPAR for providing source concentrations. The subjects' reactions to the advice also demonstrated that the user is always in control in that the advice does not have to be accepted. In the absence of a known source concentration, EXPAR first recommends the use of the chemical's solubility limit or a concentration value from a similar facility. The subjects were not able to do either because acetone is infinitely soluble (i.e., it is miscible in all proportions) and concentration data for other facilities are not included in EXPAR (acetone is not on the EPA's Priority Pollutant list). Given this situation, EXPAR

recommends that the user work with relative concentrations wherein the source is assigned a "concentration" of 100%, and computed values of 0.1% are taken to define the limit of migration for the chemical. Interestingly, none of the four subjects followed this recommendation. Test Subjects 3, 8 and 17 assigned concentration values based on unspecified reasons (assumed to be guessed values). Test Subject 10, who had concentration data for three sampling points located downgradient from the source, simply assigned a value that was larger than these. It is worth noting that one of the subjects assigned DS 4 used the relative concentration approach. The length determined for the acetone plume was 500 meters, only slightly longer than the actual length of 460 meters.

The simulated plumes obtained by Test Subjects 3, 10 and 17 for TCE are reasonable matches to the actual plume configuration. The variations in the configurations are mainly a function of differences in advection rate and in the organic carbon content value used in the calculation of a retardation factor. In the case of the latter, EXPAR recommended a default value of 0.001 to Test Subjects 3 and 10. However, Test Subject 17 was given a value (0.01) in DS 4. The range of values reported in Jackson et.al. (1985) is 0.0015 to 0.0131. Therefore, the default value of 0.001 can not be considered to be inaccurate in this particular case. With regard to the transformation of TCE, all test subjects used a half-life value of 320 days. This is the value recommended by the system for abiotic hydrolysis.

All of the simulated plume configurations for DCE are poor estimates. There are two possible explanations for this result. First, it is possible that DCE was introduced to the Compound during the

latter stages of operation. Consequently, the plume has developed over a period of time that is less than 15 years. Second, the half-life value for abiotic hydrolysis used by the subjects may not be accurate. Callahan et.al. (1979) indicate that the half-life may be in the order of 50000 years. However, they also point out that it could be as low as 6 to 18 months or, possibly, even only about 40 days. Because these investigators indicate a preference for the 50000-year value, this was the one entered into EXPAR's data base of chemical data. All of the subjects used this value in their simulations. To test whether a shorter half-life would yield better results, I reran the data set prepared by Test Subject 3 with a half-life of 40 days. The length of the resulting plume was 60 meters, which is the same as the length of the actual DCE plume. The importance of my simulation is simply that the half-life values suggested by EXPAR could be inaccurate. EXPAR does advise the user of this and recommends that half-life values for any chemical should be researched and checked before use.

Comparisons of the plume configurations derived by Test Subjects 3 and 17 provide indications as to the validity of the heuristics employed by EXPAR. Test Subject 3 utilized a significant amount of heuristic information as 18 of the 22 parameter values were obtained directly or indirectly from EXPAR (Table 6). In contrast, TS 17 obtained only 8 values from the system. Based on plume lengths, the results provided by TS 3 for acetone are more accurate than those determined by TS 17 whereas the opposite is true for TCE. However, the variations can not be considered to be large. The results for DCE provided by both subjects were approximately the same although inaccurate, as noted above. Overall, it is apparent that the plumes

Table 6 Derivation of Parameter Values

Data Set No.	Test Subject No.	Given	Number of Values: Provided by EXPAR	Provided by Subject	Provided by Subject with Advice from EXPAR	Not Determined
1	3	2	15	1	3	1
2	8	4	10	4	3	1
3	10	4	10	5	3	0
4	17	5	7	8	1	1

predicted largely on the basis of heuristic values were not significantly different relative to those prepared using the best available information. Therefore, the heuristics employed by EXPAR appear to be valid for this particular site.

In terms of risk assessment or the planning of a first-stage monitoring strategy, the results obtained by TS 10 would be satisfactory for all three chemicals modeled. For Test Subjects 3 and 17, the predictions of ACE and TCE concentrations would be acceptable but that of DCE would not. The results provided by TS 8 are obviously unsatisfactory for assessing risk and planning field studies.

The evaluation of the human engineering aspects was done on an informal basis. No system "crashes" were known to have occurred. Based on several interviews, the test subjects found the system simple to operate and easy to understand. This result was particularly significant in view of the fact that some of the subjects had little previous experience with transport models and computers. In fact, TS 17 had never used a computer prior to two weeks before the exercise was initiated. The only improvement suggested and later added was to have a means with which a partially completed set of data could be saved in the event that a session was interrupted and had to be completed at a later time. Overall, the subjects were satisfied with the performance and capabilities of the system.

The results provided by the remaining 14 test subjects were not examined in detail. However, a general assessment of plume lengths was made. This showed that the plumes determined by seven subjects were reasonable facsimiles of the actual configurations while the opposite was true for the other seven. The poor results obtained by the latter

subjects were caused in part by an inadequate understanding of hydrogeology. However, it was obvious that the time and effort given to the exercise was an important factor.

6. CONCLUSIONS

The results of this investigation have illustrated that expert systems can be developed to solve problems associated with the implementation of a transport model. The investigation also has demonstrated that the complexity of planning a monitoring strategy could be alleviated through the use of these techniques.

The knowledge representation structures employed by EXPAR are effective for storing and conveying the information required to apply the ROKEY model. Of particular importance is the incorporation of existing software that increases the power of the system. By partitioning EXPAR with a set of computer forms at the top level, expert users should find the system useful. These users likely will take advantage of several of the bottom level programs, such as the interfaces to the data bases and the consistency checking facilities. Non-expert users can apply the model to contamination problems within a short period of time. This group even would include those who are unfamiliar with models and computers. This user-friendly capability has been achieved with the assistance, advisory and tutorial facilities in the bottom level. Related to this is the most unique feature of EXPAR -- the formalization of reasoning strategies and other heuristic knowledge in the domain of contaminant hydrogeology.

At the present time EXPAR is at an advanced stage of development. The results of the evaluation exercise indicate a high level of performance both in terms of assigning valid parameter values and in human engineering considerations. However, there has been only one evaluation. A full assessment can be made only after the system has

been applied to a large number of sites and types of sources. The exercise did reveal several weaknesses in EXPAR, particularly with the assignment of hydraulic conductivity values, the validity of half-life values, and the lack of assistance in determining the principal direction of transport. Most of the knowledge currently in the system has been obtained from public sources with the remainder coming from my own heuristic judgments. The next stage of development would involve the careful scrutiny of the system by recognised experts. This would be followed by the refinement and possible expansion of the system to include improved heuristic knowledge. All of this further refinement could be done, at least in part, by others as various portions of the knowledge can be modified and expanded by the users. The parts of the system that users can modify and the instructions for doing so are given in the operations manual for Expert ROKEY (McClymont and Schwartz, 1987). Nonetheless, I believe that the current version of EXPAR could be used successfully in many contamination studies.

The numerous structures used to build EXPAR has resulted in a flexible and expressive system. One of the benefits that can come from this is the potential for applying the methodology to similar problems in the domain. In the present sense, this would involve the adaptation of EXPAR to other contaminant transport codes. Work in this area has already begun where Mattson (1986) used computer forms in a preprocessor for the USGS MOC (Method of Characteristics) transport model. The next step of adding bottom level programs would be straightforward, particularly in view of the fact that ROKEY and MOC share several common parameters. It follows that the architecture and structures of EXPAR could be applied to other transport codes or, more

generally, to any computer program that requires a large set of input data.

The current design of EXINS appears to be an effective architecture for accomplising its task of recommending a monitoring strategy. The public knowledge that it contains was conducive to representation using the production rule formalism. The architecture employed minimizes some of the problems that have been identified with traditional rule-based systems. The next stage in the development process would be to refine, expand and test the knowledge in the system. However, it is important to note that the current architecture may become unsuitable in the rule set increases to over 200. Specifically, the time to compile the program may become excessively long and the organization of the rules could become complex. Therefore, consideration would have to be given to redesigning or replacing the architecture. One possible alternative is to combine the rules with frames where a frame represents one of the strategy topics in the system.

REFERENCES

Abriola, L.M. and G.F. Pinder. 1982. Calculation of Velocity in Three Space Dimensions from Hydraulic Head Measurements. Ground Water, Vol.20, No.2, pp.205-213.

Aikins, J.S. 1983. Prototypical Knowledge for Expert Systems. Artificial Intelligence, Vol.20, No.2, pp.163-210.

Aikins, J.S. 1984. A Representation Scheme Using Both Frames and Rules. In: Rule-Based Expert Systems, Chapter 23, pp.424-440. eds. B.G. Buchanan and E.H. Shortliffe, Addison-Wesley Publishing Co. 748 pp.

Aller, L., T. Bennett, J.H.Lehr and R.J. Petty. 1985. DRASTIC: A Standardized System for Evaluating Ground Water Pollution Potential Using Hydrogeologic Settings. U.S. EPA Report No. EPA/600/2-85/018. 163 pp.

Alty, J.L. and M.J. Coombs. 1984. Expert Systems Concepts and Examples. NCC Publications, Mancester, England, 209 pp.

Barcelona, M.J., J.P. Gibbs and R.A. Miller. 1984. A Guide to the Selection of Materials for Monitoring Well Construction and Ground-Water Sampling. U.S. EPA Report No. EPA-600/2-84-024. 78 pp.

Buchanan, B.G. and E.H. Shortliffe. 1984. Human Engineering of Medical Expert Systems. In: Rule-Based Expert Systems, Chapter 32, pp.599-612. eds. B.G. Buchanan and E.H. Shortliffe, Addison-Wesley Publishing Co. 748 pp.

Callahan, M.J., M.W. Slimak, N.W. Gobel, I.P. May, C.F. Fowler, J.R.Freed, P. Jennings, R.L. Durfee, F.C. Whitmore, B. Maestri, W.R. Mabey, B.R. Holt and C. Gould. 1979. Water-Related Fate of 129 Priority Pollutants. U.S. EPA Report No. EPA-440/4-79-029b. 673 pp.

Canter, L.W. 1985. Methods for the Assessment of Ground Water Pollution Potential. In: Ground Water Quality, Chapter 13, pp.270-306. eds. C.H. Ward, W. Giger and P.L. McCarty, John Wiley and Sons. 547 pp.

Cherry, J.A., R.W. Gillham, E.G. Anderson and P.E. Johnson. 1983. Migration of Contaminants in Groundwater at a Landfill: A Case Study. 2. Groundwater Monitoring Devices. J. Hydrol., Vol.63, pp.31-49.

Domenico, P.A. 1987. An Analytical Model for Multidimensional Transport of a Decaying Contaminant Species. J. Hydrol., Vol.91, pp.49-58.

Domenico, P.A. and G.A. Robbins. 1985. A New Method of Contaminant Plume Analysis. Ground Water, Vol.23, No.4, pp.476-485.

Donigian, A., T.Y.R. Lo and E. Shanahan. 1983. Rapid Assessment of Potential Ground-Water Contamination Under Emergency Response Conditions. U.S. EPA Report No. EPA-600/8-83-030. 147 pp.

Federal Register. 1980. Summary of USEPA Priority Pollutants. Vol.45, November 28, 1980, pp.79324-79341.

Fenn, D., E. Cocozza, J. Isbister, O. Braids, B. Yare and P. Roux. 1977. Procedures Manual for Ground Water Monitoring at Solid Waste Disposal Facilities. U.S. EPA Report No. EPA/SW-611. 269 pp.

Goodall, A. 1985. The Guide to Expert Systems. Learned Information (Europe) Ltd., Abingdon, England. 220 pp.

Graham, A.J. 1985. Gloucester Landfill Waste Site Engineering Study. Problem Definition and Remedial Alternatives. Unpublished consultants report prepared for Transport Canada, Project No. L10137-1174, by A.J. Graham Engineering Consultants Ltd., Water and Earth Science Associates Ltd. and Canviro Consultants Ltd.

Hart, P. 1982. Directions for AI in the Eighties. SIGART Newsletter, Vol.79, pp.11-16.

Jackson, R.E., R.J. Patterson, B.W. Graham, J. Bahr, D.W. Belanger, J. Lockwood and M. Priddle. 1985. Contaminant Hydrogeology of Toxic Organic Chemicals at a Disposal Site, Gloucester, Ontario: 1. Chemical Concepts and Site Assessment. IWD Scientific Series No. 141, Environment Canada, Ottawa, Ontario. 114 pp.

Mattson, B. 1986. MOCIN. Unpublished research paper, Dept. of Geology, Univ. of Alberta.

McClymont, G.L. and F.W. Schwartz. 1987. Development and Application of an Expert System in Contaminant Hydrogeology. The Expert ROKEY Computer System. Final Report and Users Manual. Unpublished consultants report prepared for the National Hydrology Research Institute, Environment Canada by Simco Groundwater Research Ltd.

Milne-Home, W.A. 1985. Modelling Mass Transport in Clastic Sands and Sandstones. Ph.D Thesis, Univ. of Alberta, Edmonton, Alberta, Canada.

Negoita, C.V. 1985. Expert Systems and Fuzzy Systems. Benjamin/Cummings Publishing Co., Menlo Park, Calif. 190 pp.

Pettyjohn, W.A., W.J. Dunlap, R.L. Cosby and J.W. Keely. 1981. Sampling Ground Water for Organic Contaminants. Ground Water, Vol.19, No.2, pp.180-189.

Sanders, T.G., R.C. Ward, J.C. Loftis, T.D. Steele, D.D. Adrian and V. Yevjevich. 1983. Design of Networks for Monitoring Water Quality. Water Resources Publications, Littleton, Col., 328 pp.

Schwartz, F.W. and A.S. Crowe. 1980. A Deterministic-Probabilistic Model for Contaminant Transport. Consultants report prepared for U.S. Nuclear Regulatory Commission by CGS, Inc. NUREG/CR-1609, CGS/NR85U060. 158 pp.

Schwartz, F.W., G.L. McClymont and L. Smith. 1985. On the Role of Mass Transport Modeling. In: Hazardous Wastes in Ground Water: A Soluble Dilemma. pp.2-12. Proceedings Second Canadian/American Conference on Hydrogeology, Banff, Alberta, Canada, June 25-29, 1985. eds. B. Hitchon and M.R. Trudell. National Water Well Association, Dublin, Ohio. 255 pp.

Silverman, B.G. 1986. Reflections on Some Next Generation of AI Tools. In: Expert Systems in Government Symposium, pp.427-427. eds. K.N. Karna, K. Parsaye and B.G. Silverman. IEEE Computer Society Press, Washington, D.C. 466 pp.

Waterman, D.A. 1986. A Guide to Expert Systems. Addison-Wesley Publishing Co., Reading, Mass. 419 pp.

Weiss, S.M. and C.A. Kulikowski. 1984. A Practical Guide to Designing Expert Systems. Rowman and Allanheld Publishers, Totowa, N.J. 174 pp.

APPENDIX A

References Cited in Expert ROKEY

Abriola, L.M. and G.F. Pinder. 1982. Calculation of Velocity in Three Space Dimensions from Hydraulic Head Measurements. Ground Water, Vol.20, No.2, pp.205-213.

Anderson, M.P. 1979. Using Models to Simulate the Movement of Contaminants Through Groundwater Flow Systems. Crit.Rev.Environ.Controls, Vol.9, No.2, pp.97-156.

Anderson, M.P. 1984. Movement of Contaminants in Groundwater: Groundwater Transport - Advection and Dispersion. In: Studies in Geophysics, Groundwater Contamination. Panel on Groundwater Contamination. National Academy Press. Washington, D.C.

Barcelona, M.J., J.P. Gibbs and R.A. Miller. 1984. A Guide to the Selection of Materials for Monitoring Well Construction and Ground-Water Sampling. U.S. EPA Report No. EPA-600/2-84-024.

Bear, J. 1972. Dynamics of Fluids in Porous Media. American-Elsevier, N.Y., 764 pp.

Callahan, M.J., M.W. Slimak, N.W. Gobel, I.P. May, C.F. Fowler, J.R. Freed, P. Jennings, R.L. Durfee, F.C. Whitmore, B. Maestri, W.R. Mabey, B.R. Holt and C. Gould. 1979. Water-Related Fate of 129 Priority Pollutants. U.S. EPA, EPA-440/4-79-029b.

Carter, C.W. and I.H. Suffet. 1982. Binding of DDT to Dissolved Humic Materials. Environ. Sci. Technol., Vol.16, No.11, pp.735-740.

Cherry, J.A., R.W. Gillham, E.G. Anderson and P.E. Johnson. 1983. Migration of Contaminants in Groundwater at a Landfill: A Case Study. 2. Groundwater Monitoring Devices. J. Hydrol., Vol.63, pp.31-49.

Cherry. J.A., R.W. Gillham and J.F. Barker. 1984. Contaminants in Groundwater: Chemical Processes. In: Studies in Geophysics, Groundwater Contamination. Panel on Groundwater Contamination. National Academy Press, Washington, D.C.

Chiou, C.T., L.J. Peters and V.H. Freed. 1979. A Physical Concept of Soil-Water Equilibria for Nonionic Organic Compounds. Science, Vol.206, pp.831-832.

Chiou, C.T., P.E. Porter and D.W. Schmedding. 1983. Partition Equilibria of Nonionic Organic Compounds Between Soil Organic Matter and Water. Environ. Sci. Technol., Vol.17, No.4, pp.227-231.

Dagan, G. 1982. Stochastic Modeling of Groundwater By Unconditional and Conditional Probabilities, 2. The Solute Transport. Wat. Resour. Res., Vol.18, pp.835-848.

Domenico, P.A. and G.A. Robbins. 1985. A New Method of Contaminant Plume Analysis. Ground Water, Vol.23, No.4, pp.476-485.

Egboka, B.C.E., J.A. Cherry, R.N. Farvolden and E.O. Frind. 1983. Migration of Contaminants in Groundwater at a Landfill: A Case Study. 3. Tritium as an Indicator of Dispersion and Recharge. J. Hydrol., Vol.63, pp.51-80.

Ellis, J.H., R.I. Barnhisel and R.E. Phillips. 1970. The Diffusion of Copper, Manganese, and Zinc as Affected by Concentration, Clay Mineralogy, and Associated Anions. Soil Sci. Soc. Amer. Proc., Vol.34, pp.866-870.

Fair, G.M. and L.P. Hatch. 1933. Fundamental Factors Governing the Streamline Flow of Water Through Sand. J. Amer. Water Works Assoc., Vol.25, No.11, pp.1551-1565.

Federal Register. 1980. Summary of USEPA Priority Pollutants. Vol.45, November 28, 1980, pp.79324-79341.

Fenn, D., E. Cocozza, J. Isbister, O. Braids, B. Yare and P. Roux. 1977. Procedures Manual for Ground Water Monitoring at Solid Waste Disposal Facilities. U.S. EPA Report No. EPA/SW-611.

Freeze, R.A. and J.A. Cherry. 1979. Groundwater. Prentice-Hall, Englewood Cliffs, N.J., 604 pp.

Freyberg, D.L. 1986. A Natural Gradient Experiment on Solute Transport in a Sand Aquifer, 2. Spatial Moments and the Advection and Dispersion of Nonreactive Tracers. Wat. Resour. Res., Vol.22, No.13, pp.2031-2046.

Fried, J.J. 1975. Groundwater Pollution. Elsevier, Amsterdam, 330 pp.

Gelhar, L.W., A.L. Gutjahr and F.L. Naff. 1979. Stochastic Analysis of Macrodispersion in a Stratified Aquifer. Wat. Resour. Res., Vol.15, pp.1387-1397.

Gelhar, L.W. and C.L. Axness. 1981. Stochastic Analysis of Macrodispersion in Three-Dimensionally Heterogeneous Aquifers. Geophysical Research Center, Hydrology Research Program, Rep. No. H8, New Mexico Institute of Mining and Technology, Socorro, New Mexico, 140 pp.

Gelhar, L.W. and C.L. Axness. 1983. Three-dimensional Stochastic Analysis of Macro-Dispersion in Aquifers. Wat. Resour. Res., Vol.19, pp.161-180.

Gelhar, L.W., A. Mantoglou, C. Welty and K.R. Rehfeldt. 1985. A Review of Field-Scale Physical Solute Transport Processes in Saturated and Unsaturated Porous Media. Final Report EA-4190, Electric Power Research Institute, Palo Alto, California.

Goodall, D.C. and R.M. Quigley. 1977. Pollutant Migration from Two Sanitary Landfill Sites near Sarnia, Ontario. Can. Geotech. J., Vol.14, pp.223-236.

Griffin, R.A., R.E. Hughes, L.R.Folmer, C.J. Stohr, W.J. Morse, T.M. Johnson, J.K. Bartz, J.D. Steele, K. Cartwright, M.M. Killey, and P.D. DuMontelle. 1984. Migration of Industrial Chemicals and Soil-Waste Interactions at Wilsonville, Illinois. In: Proceedings of the 10th Annual Research Symposium of the Solid and Hazardous Waste Research Division, U.S. EPA, Cincinnati, Ohio.

Grisak, G.E., J.F. Pickens and J.A. Cherry. 1980. Solute Transport Through Fractured Media: 2. Column Study of Fractured Till. Wat. Resour. Res., Vol.16, pp.731-739.

Grisak, G.E. and J.A. Pickens. 1981. An Analytical Solution for Solute Transport Through Fractured Media with Matrix Diffusion. J. Hydrol., Vol.52, pp.47-57.

Grove, D.B. and W.A. Beetum. 1971. Porosity and Dispersion Constant Calculations for a Fractured Carbonate Aquifer Using the Two-Well Tracer Method. Wat. Resour. Res., Vol.7, pp.128-134.

Gupta, S.K., K.K. Tanji and J.N. Luthin. 1975. A Three-Dimensional Finite Element Groundwater Model. California Water Research Center, Contribution 152, 119 pp.

Harris, J.C. 1982. Rate of Hydrolysis. Chapter 7 in Lyman et.al. (1982), (see reference below).

Jackson, R.E., R.J. Patterson, B.W. Graham, J. Bahr, D.W. Belanger, J. Lockwood and M. Priddle. 1985. Contaminant Hydrogeology of Toxic Organic Chemicals at a Disposal Site, Gloucester, Ontario: 1. Chemical Concepts and Site Assessment. IWD Scientific Series No. 141, Environment Canada, Ottawa, Ontario. 114 pp.

Karickhoff, S.W., D.S. Brown and T.A. Scott. 1979. Sorption of Hydrophobic Pollutants on Natural Sediments. Water Res., Vol.13, pp.241-248.

Karickhoff, S.W. 1981. Semi-Empirical Estimation of Sorption of Hydrophobic Pollutants on Natural Sediments and Soils. Chemosphere, Vol.10, No.8, pp.833-846.

Klotz, D., K.P. Seiler, H. Moser and F. Neumaier. 1980. Dispersivity and Velocity Relationships From Lab and Field Experiments. J. Hydrol., Vol.45, pp.169-184.

Lallemand-Barres, A. and P. Peaudecerf. 1978. Recherche des Relations Entre la Valeur de la Dispersivite Macroscopique d'un Milieu Aquifere, ses Autres Caracteristiques et les Conditions de Mesure. Bulletin du B.R.G.M. (deuxieme serie), Section III, No.4, pp.277-284.

LaVenue, A.M. and P.A. Domenico. 1986. A Preliminary Assessment of the Regional Dispersivity of Selected Basalt Flows at the Hanford Site, Washington, U.S.A. J. Hydrol., Vol.85, pp.151-167.

Lyon, K. 1985. Permeability and Grain Size Curves. Wat. Well J., Vol.39, No.2, pp.28-29.

Lyman, W.J. 1982. Adsorption Coefficient for Soils and Sediments. Chapter 4 in Lyman et.al. (1982) (see reference below).

Lyman, W.J., W.F. Reehl and D.H. Rosenblatt. 1982. Handbook of Chemical Property Estimation Methods. Environmental Behavior of Organic Compounds. McGraw-Hill Book Co., NY.

MacFarlane, D.S., J.A. Cherry, R.W. Gillham and E.A. Sudicky. 1983. Migration of Contaminants in Groundwater at a Landfill: A Case Study. 1. Groundwater Flow and Plume Delineation. J. Hydrol., Vol.63, pp.1-29.

Masch, F.D. and K.J. Denny. 1966. Grain Size Distribution and Its Effects on the Permeability of Unconsolidated Sands. Wat. Resour. Res., Vol.2, No.4, pp.665-677.

Matheron, G. and G. DeMarsily. 1980. Is Transport in Porous Media Always Diffusive? Wat. Resour. Res., Vol.16, pp.901-917.

McCarty, P.L., M. Reinhard and B.E. Rittman. 1984. Trace Organics in Groundwater. Environ. Sci. Technol., Vol.15, No.1, pp.40-51.

McGraw-Hill. 1984. McGraw-Hill Dictionary of Chemical Terms. McGraw-Hill Book Co., NY.

Mercado, A. 1967. The Spreading Pattern of Injected Water in a Permeability Stratified Aquifer. In: Proc. Int. Assoc. Sci. Hydrol. Symp., Haifa. Publication No.72, pp.23-26.

Miller, D.W. (ed.). 1980. Waste Disposal Effects on Ground Water. Premier Press, Berkeley, Calif.

Milne-Home, W.A. 1985. Modelling Mass Transport in Clastic Sands and Sandstones. Ph.D Thesis, University of Alberta, Edmonton, Alberta, Canada.

Pettyjohn, W.A., W.J. Dunlap, R.L. Cosby and J.W. Keeley. 1981. Sampling Ground Water for Organic Contaminants. Ground Water, Vol.19, No.2, pp.180-189.

Pickens, J.F. and G.E. Grisak. 1981a. Scale-Dependent Dispersion in a Stratified Aquifer. Wat. Resour. Res., Vol.17, pp.1191-1212.

Pickens, J.F. and G.E. Grisak. 1981b. Modeling of Scale-Dependent Dispersion in Hydrogeological Systems. Wat. Resour. Res., Vol.17, pp.1701-1711.

Pinder, G.F., M. Celia and W.G. Gray. 1981. Velocity Calculations from Randomly Located Hydraulic Heads. Ground Water, Vol.19, No.3, pp.262-264.

Rao, P.S.C., J.M. Davidson, V.E. Berkheiser, L.T. Ou, J.J. Street, W.B. Wheeler and T.L. Yuan. 1982. Retention and Transformation of Selected Pesticides and Phosphorous in Soil-Water Systems: A Critical Review. U.S. EPA, EPA-600/3-82-060, Athens, Georgia.

Robson, S.G. 1974. Feasibility of Digital Water-Quality Modeling Illustrated By Application at Barstow, Ca. U.S. Geological Survey Water Research Investigation Report 46-73, 66 pp.

Sabel, G.V. and T.P. Clark. 1984. Volatile Organic Compounds as Indicators of Municipal Solid Waste Leachate Contamination. Waste Man. Res., Vol.2, pp.119-130.

Saxena, S.K., L. Boersma, F.T. Lindstrom and J.L. Young. 1974a. The Self-Diffusion Coefficients of 45Ca and 2,4-Dichlorophenoxyacetic Acid. Soil Sci., Vol.117, No.1, pp.14-20.

Saxena, S.K., L. Boersma, F.T. Lindstrom and J.L. Young. 1974b. Effect of Pore Size on Diffusion Coefficients in Porous Media. Soil Sci., Vol.117, No.1, pp.80-86.

Schwartz, F.W. 1977. On Radioactive Waste Management: Model Analysis of a Proposed Site. J. Hydrol., Vol.32, pp.257-277.

Schwarzenbach, R.P. and J. Westall. 1981. Transport of Nonpolar Organic Compounds from Surface Water to Groundwater. Laboratory Sorption Studies. Environ. Sci. Technol., Vol.15, No.11, pp.1360-1367.

Scow, K.M. 1982. Rate of Biodegradation. Chapter 9 in Lyman et.al. (1982), (see reference above).

Smith, L. and F.W. Schwartz. 1980. Mass Transport. 1. A Stochastic Analysis of Macroscopic Dispersion. Wat. Resour. Res., Vol.16, pp.303-313.

Smith, L. and F.W. Schwartz. 1981a. Mass Transport. 2. Analysis of Uncertainty in Prediction. Wat. Resour. Res., Vol.17, pp.351-369.

Smith, L. and F.W. Schwartz. 1981b. Mass Transport. 3. Role of Hydraulic Conductivity Data in Prediction. Wat. Resour. Res., Vol.17, pp.1463-1479.

Sudicky, E.A., J.A. Cherry and E.O. Frind. 1983. Migration of Contaminants in Groundwater at a Landfill: A Case Study. 4. A Natural-Gradient Dispersion Test. J. Hydrol., Vol.63, pp.81-108.

Sutton, P.A. and J.F. Barker. 1985. Migration and Attenuation of Selected Organics in a Sandy Aquifer. Ground Water, Vol.23, No.1, pp.10-16.

van Schaik, J.C., W.D. Kemper and S.R. Olsen. 1966. Contribution of Adsorbed Cations to Diffusion in Clay-Water Systems. Soil Sci. Soc. Amer. Proc., Vol.30, pp.17-25.

Walton, W.C. 1970. Groundwater Resource Evaluation. McGraw-Hill Book Co., NY.

Walton, W.C. 1984. Practical Aspects of Ground Water Modeling. National Water Well Association.

APPENDIX B

EXPAR Computer Forms

Form 1

You may find it useful to fill in the information below to help keep track of this session. All entries, however, are optional.

Project no.: _____ Date (mo/day/yr): _____

Project name: _____ Investigator: _____

Site name: _____ Name of stratum: _____

Comments: _____

_____

_____

_____

_____

F1:back 2:next 3:enter 4:pg? 5:elab 6:assist 7:pg.bk 8:pg.fwd 9:--- 10:exit

Form 2

Before continuing, we need entries for the five parameters listed below. This information will be used for consistency checking and providing assistance later in the session. If you want an elaboration of any parameter, just move the cursor beside the parameter and press F5. To derive the value of a parameter, press F6 (do not press F3).

Name of chemical:

Type of source:

Type of stratum material:

Slope of ground surface:

Principal direction of horizontal ground water flow:

F1:back 2:next 3:enter 4:pg? 5:elab 6:assist 7:pg.bk 8:pg.fwd 9:--- 10:exit

Form 3

The following transport and simulation information is required:

Advection rate (cm/s): _____

Horizontal hydraulic gradient (fraction): _____

Effective porosity (fraction): _____

Dispersivity - longitudinal (x): _____

(cm) - transverse-horizontal (y): _____

- transverse-vertical (z): _____

Effective diffusion coeff. (cm2/s): _____

Simulation period: _____ days or _____ years

F1:back 2:next 3:enter 4:pg? 5:elab 6:assist 7:pg.bk 8:pg.fwd 9:--- 10:exit

Form 4

The following information related to the size of the source is required. Depth refers to distance below ground surface.

Length of source (meters): _____

Depth to top of source (meters): _____

Depth to bottom of source (meters): _____

Depth to water table (meters): _____

Depth to base of stratum (meters): _____

F1:back 2:next 3:enter 4:pg? 5:elab 6:assist 7:pg.bk 8:pg.fwd 9:--- 10:exit

Form 5

The following time and concentration data are needed. ROKEY works in relative time so the initial entry of the chemical into the ground water system is set at t=0 days. All other times (if applicable) are specified relative to this. Any time less than 100,000 days can be entered.

Solubility: _____ Units: _____

Initial concentration: _____ Units: _____

Time Concentration Changed (days)	New Concentration
_____	_____
_____	_____
_____	_____
_____	_____

(continue on next form if needed)

F1:back 2:next 3:enter 4:pg? 5:elab 6:assist 7:pg.bk 8:pg.fwd 9:--- 10:exit

Form 6

Time Concentration Changed (days)	New Concentration

F1:back 2:next 3:enter 4:pg? 5:elab 6:assist 7:pg.bk 8:pg.fwd 9:--- 10:exit

Form 7

Values for the parameters listed below are needed if retardation via sorption is to be accounted for. If you do not want to consider sorption, just enter a value of 0 for "Organic carbon content" and leave the other slots blank.

Organic carbon content (fraction): _____

Organic carbon / water partition coeff. (log value): _____

Total porosity (fraction): _____

Dry bulk density (g/cm3): ____

F1:back 2:next 3:enter 4:pg? 5:elab 6:assist 7:pg.bk 8:pg.fwd 9:--- 10:exit

Form 8

The transformation or decomposition of the chemical due to abiotic hydrolysis or biotransformation can be accounted for by providing a half-life value. The program has been set-up to use one value only and, as such, the smallest one entered that is greater than zero will be used by ROKEY.

Hydrolytic (abiotic) half-life: _____ days or _____ years

Biotransformation half-lives:

Aerobic conditions - _____ days or _____ years

Anaerobic conditions - _____ days or _____ years

F1:back 2:next 3:enter 4:pg? 5:elab 6:assist 7:pg.bk 8:pg.fwd 9:--- 10:exit

Form 9

Before specifying the output data and running ROKEY, do you want :

a) the data set evaluated ? (F5= elab, F6= yes) :

b) a check of the plume length ? (F5= elab, F6= yes) :

Please enter the following output data:

X interval (meters): _____

XY Plot: Y interval (meters): _____
Z coordinate of XY plane: __
XZ Plot: Y coordinate of XZ plane: __

Name to be given to this input file: _____
Name to be given to the output file: _____

Name to be given to the plot file: _____

Do you want a listing of the data set in the output file ? (Y/N) : _

F1:back 2:next 3:enter 4:pg? 5:elab 6:assist 7:pg.bk 8:pg.fwd 9:--- 10:exit

APPENDIX C

Principal Tutorial Topics in Elaboration Programs

Transport Process	Parameter That Tutorial is Accessed From	Tutorial Topics
Advection	Advection rate	Summary
Mechanical dispersion	Dispersivity - longitudinal (x)	Level 1 Dispersivity and Dispersion in Contaminant Transport Description of the Dispersion Process Dispersion is Directional Microscopic Dispersion Macroscopic Dispersion Why Dispersion is Important Definition of Dispersivity Deriving Dispersivity Values Values of Longitudinal Dispersivity Level 2 Fick's Second Law Sites With Fickian Dispersion Sites Without Fickian Dispersion New Approaches in Transport Modeling
Diffusion	Effective diffusion coeff.	Level 1 Definition of Diffusion Diffusion in Contaminant Transport Values for the Effective Diffusion Coefficient Level 2 Sites Where Diffusion Could Be Important Molecular Diffusion Coeffs. Effective Diffusion Coeffs. Apparent Diffusion Coeffs.
Sorption	Organic carbon/water partition coeff.	Summary Sorption Estimating Koc Values Organic Carbon Content Reliability of Retardation Factors

Transport Process	Parameter That Tutorial is Accessed From	Tutorial Topics
Abiotic Hydrolysis	Hydrolytic (abiotic) half-life	Introduction Rate of Hydrolysis Products of Hydrolysis Half-life Values in This Computer System
Biotransformation	Aerobic conditions or Anaerobic conditions	Introduction The Process of Biotransformation Factors Effecting Biotransformation Rate of Biotransformation Half-life Values in This Computer System

APPENDIX D

Summary of Contents of Assistance Programs

Parameter	Type of Assistance
Form 1 bookkeeping parameters	- no assistance provided; entries are volunteered by the user
Name of chemical	- user selects from menu in CHEM.LST; currently contains 118 organic chemicals and isomers from U.S. EPA 129 Priority Pollutant List - transfers plausible values of log Kow and solubility from CHEM.DAT to the global data base - transfers default values of abiotic and biotransformation half-lives from CHEM.DAT to the global data base
Type of source	- user selects from menu - menu lists 12 types of municipal, domestic and industrial sources - access is given to lists of types of chemicals (and reported concentrations) associated with each source
Type of stratum material	- user selects from menu - menu lists 12 material types - transfers default values of hydraulic conductivity, porosity, dry bulk density and organic carbon content from STRAT.DAT to the global data base
Slope of ground surface	- user selects from menu - menu lists 9 choices ranging from 0.01% to 30% - default hydraulic gradient derived from rules using slope and type of stratum material; the value is entered into the global data base
Principal direction of horizontal ground water flow	- advice provided in a text file
Advection rate	- user is interviewed for data with which a value can be assigned or calculated - questioning sequence designed to derive value from most-preferred to least-preferred methods; terminates with the assignment of value derived using values of effective porosity, hydraulic gradient and default value of hydraulic conductivity (determined in "Type of stratum material") - principal structures are a control block, a Q/A network and subprograms - two data files available (VEL.DAT and K.DAT) - program contains existing codes for calculating hydraulic conductivity from grain-size data; the codes include GENPAR1 (Milne-Home, 1985) and codes that solve the Kozeny-Carman and Hazen equations;

Parameter	Type of Assistance
Advection rate (cont.)	these codes are accessed from Questions 12, 13 and 14, respectively; all codes are adapted for interactive usage - user can check his volunteered value with a default value calculated by the program
Horizontal hydraulic gradient	- user is interviewed for data with which a value can be assigned or calculated - questioning sequence designed to derive value from most-preferred down to least-preferred methods; terminates with the assignment of a default value determined in "Slope of ground surface" - principal structures are a control block, a Q/A network and subprograms - one data file is available (I.DAT) - program contains a conventional code that calculates a value using a linear interpolation scheme (Abriola and Pinder, 1982); the code operates interactively and is accessed from Q.2 - user can check his volunteered value with the default value described above
Effective porosity	- user can select a value from a data file (POR.DAT) or accept the default value determined in "Type of stratum material" - principal structures are a control block, a Q/A network and a data file - user can check value selected from data file with default value
Dispersivity - longitudinal (x)	- user is interviewed for data with which a value can be assigned - questioning sequence designed to derive value from most-preferred to least-preferred method; terminates with the assignment of a default value - principal structures are a control block, a Q/A sequence and a rule set - one data file is available (DX.DAT) - rules used to define the scale of transport - rules determine initial questioning order - rules assign default value; default value is 0 if advection rate is less than 10-5 cm/s or mean travel distance of plume is less than 0.1 meters
Dispersivity - transverse-horizontal (y)	- rules assign default value - default value chosen on basis of longitudinal value
Dispersivity - transverse-vertical (z)	- rules assign default value - default value chosen on basis of longitudinal value

Parameter	Type of Assistance
Effective diffusion coeff.	- rules assign default value - default value chosen on basis of type of stratum material
Simulation period	- no assistance provided
Length of source	- advice given in a text file - advice based on "Type of source"
Depth to top of source	- same as above
Depth to bottom of source	- same as above
Depth to water table	- advice given in a text file - advice based on general assumptions
Depth to base of stratum	- same as above
Solubility	- a value, if available in CHEM.DAT, is automatically placed in the slot - if a value is not available, user is given a list of references in a text file
Initial concentration	- advice given in a text file
Changes in concentration with time	- no assistance provided
Organic carbon content	- user can select a value from a data file (FOC.DAT) or accept the default value determined in "Type of stratum material" - principal structures are a control block, a Q/A network and the data file - rules set value to 0 if the stratum material is a fractured clay - user enters 0 if sorption is not to be accounted for
Organic carbon/water partition coeff.	- calculated using Kow value from CHEM.DAT and regression equation coeff. pair selected by the user - Kow (log value) from CHEM.DAT assigned when "Name of chemical" has been selected from menu; references to literature given if value is not available - rules used to assist in selecting regression equation coeff. pair - if chemical is not in CHEM.DAT, advice for selecting regression coeff. pair is given in a text file

Parameter	Type of Assistance
Organic carbon/water partition coeff. (cont.)	- note: if user is modeling the distribution of a charged constituent (eg. heavy metals), information on entering a Kd value rather than a Koc value is given at the end of the text file KOC1.ELB, which is accessed in the explanation file
Total porosity	- user advised in a text file to use the value determined for "Effective porosity" on Form 3
Dry bulk density	- calculated automatically using an equation
Hydrolytic (abiotic) half-life	- a default value read from CHEM.DAT after "Name of chemical" has been selected from menu will be given, if available - if a default value is not available, references to the literature are given in a text file
Biotransformation half-lives -- aerobic and anaerobic conditions	- same as above
Data set evaluation (subtask)	- a rule set is used to check the consistency of the values entered on Forms 2 to 8 - any value determined to be inconsistent is reported to the user; user has the option of either changing the value or keeping it
Check of plume length (subtask)	- a variation of the ROKEY equation is used to calculate the distance from the source, along the centerline of the plume, to a user-specified concentration - to assist in selecting the target concentration, any recommended maximum concentration data for the chemical will be retrieved from LIMITS.DAT and reported to the user; the concentration data include the recommended maximum limits for human consumption, limits based on cancer risk, and taste and odor concentrations as determined by the U.S. EPA (Federal Register, November 28, 1980)
X interval	- an appropriate value is calculated on the basis of the results of "Check of plume length" - advice also is provided
Y interval	- advice based on the X interval is provided
Z coordinate of XY plane	- plane with highest concentrations either is determined with rules or is calculated - advice is provided in a text file

Parameter	Type of Assistance
Y coordinate of XZ plane	- advice in a text file is presented
Name of input file	- advice is given in a text file
Name of output file	- same as above
Name of plot file	- same as above
Listing of data set	- same as above

APPENDIX E

Strategy File (K.BAS) in EXINS

```
000 File: K.BAS
000
000                Strategy File for Monitoring Program System
000
001We suggest the following for your detection monitoring program:
001
001
002We suggest the following for your "detailed" detection monitoring program:
002
002
003We suggest the following for your interpretive monitoring program:
003
003
000 The following info. for 004 and 005 apply to all monitoring programs.
000
0041) Types of Monitoring Wells
004   -------------------------
004
004   Recommended: water-table wells and normal (single-level) piezometers
004
004   Not Recommended: multilevel piezometers (unless it is known that all
004                    of the geologic materials to be encountered are
004                    cohesionless)
004
005   Existing water-supply wells or existing monitoring wells should be used
005     only to detect the presence of a chemical of interest. Concentration
005     values or the apparent absence of a chemical should be treated with
005     caution.
005
0062) Locations of Monitoring-Well Nests
006   ----------------------------------
006
000 The info. for 007 is for a minimum detection program
000
007   The suggested configuration is:
007
007        Number of A nests - 1
007        Number of B nests - 1
007        Number of C nests - 3
007
000 The info. for 008 is for a "detailed" detection program
000
008   The suggested configuration is:
008
008        Number of A nests - 4
008        Number of B nests - 3
008        Number of C nests - 3
008
```

000 The info. for 009 is for an interpretive program
000
009 The suggested configuration is:
009
009 Number of A nests - 4
009 Number of B nests - 3
009 Number of C nests - 10
009
0103) Completion Depths of Monitoring Wells
010 ---
010
000The objective is to have 3 piezometers placed below the base of the source
000at each nest. The info. for 013 is for a source above the water table. The
000info. for 014 to 018 is for bases of sources below the water table. The
000info. in 019 is given in all cases, 020 is given if the questioning deter-
000mines that an aquifer is situated within 30 m of the water table, and 021
000is given if the unsaturated zone is 3 m thick or greater.
000
011 The suggested configuration at each nest is:
011
012 In an interpretive program, completion depths are determined during
012 test drilling. Until this commences, the suggested configuration at
012 each nest is:
012
013 One water-table well screened 1 m above and 1 m below the position of
013 the water table at the time of drilling.
013
014 Three piezometers with the screens positioned continuously with
014 depth, beginning at 1 m below the water table.
014
015 Four piezometers with the screens positioned continuously with
015 depth, beginning at 1 m below the water table.
015
016 Five piezometers with the screens positioned continuously with
016 depth, beginning at 1 m below the water table.
016
017 Six piezometers with the screens positioned continuously with
017 depth, beginning at 1 m below the water table.
017
018 Six piezometers with the screen of the deepest one placed at a
018 depth of 3 m below the base of the source. The remaining five
018 will be placed in the coarsest-grained geologic material
018 encountered during test drilling.
018
019 Additional water-table wells must be completed in "perched" water
019 tables encountered at the site of each nest.
019
020 Two or three pressure-vacuum lysimeters (or a similar device) must be
020 installed below the source within the unsaturated zone.
020

021 Piezometers should be completed in the lower aquifer:
021 - one piezometer at one of the A nests
021 - one piezometer at one of the B nests
021 - one piezometer at one of the C nests
021
0224) Monitoring Well Materials
022 ----------
022
000 Info. in 023 applies to landfill leachate, spills or leaks of crude oil
000 and derivatives, and any other water that has mod. to high concs. of
000 organic chemicals. Info. in 024 applies to water that is weakly acidic,
000 low-mod. dissolved solids content and low concs. of organic chemicals
000 such as sewage lagoon water. Info. in 025 applies to water that is either
000 acidic or has a high dissolved solids content.
000
023 Recommended: Teflon or stainless steel
023
023 Not Recommended: PVC, metals other than stainless steel (eg. iron,
023 conventional steel, galvanized steel)
023
024 Recommended: Teflon, stainless steel or PVC
024
024 Not Recommended: metals other than stainless steel (eg. iron, conven-
024 tional steel, galvanized steel)
024
025 Recommended: Teflon, PVC
025
025 Not Recommended: any metal including stainless steel, conventional steel,
025 galvanized steel and iron
025
0265) Drilling Method
026 ----------
026
000 Info. in 027 given if cobbles and boulders will be encountered. Info. in
000 028 applies to drilling and well completion in cohesive deposits and 029
000 applies to cohesionless deposits. Info. in 030 given if drilling in bed-
000 rock is anticipated.
000
027 Recommended: air rotary (with downhole hammer) or cable tool
027
028 Recommended: solid or hollow stem auger
028
029 Recommended: hollow stem auger
029
030 Recommended: air rotary or cable tool (for drilling in bedrock)
030
0316) Well Completion Reminders
031 ----------
031
000 Info. in 032, 033, 034 and 037 apply to all monitoring wells. Info. in 035
000 applies to water with high acidity or mod. to high concentrations of
000 organic chemicals; info. in 036 applies to remaining types of water
000

```
032   - complete only one well per borehole
033   - maximum screen length for piezometers is 1 m
034   - sand pack material (if required) must be washed
035   - recommended seal material: expanding, chemically-resistant cement
036   - recommended seal material: bentonite (pellets or slurry)
037   - recommended method of developing: pumping or "gentle" bailing
037
0387) Ground Water Sampling Devices
038   ---------------------------
038
000 Info. in 039 given for sites where organic chemicals are present. Info.
000 in 040 given for sites where inorganic chemicals only are present.
000
039   Recommended: positive displacement pump (eg. bladder pump);
039                flow-through cell for on-site measurement of pH, specific
039                conductance, temperature, dissolved oxygen and Eh (or pe)
039
039   Suggested: syringe sampler for volatile organics
039
039   Not Recommended: bailers
039
040   Recommended: positive displacement pump (eg. bladder pump);
040                flow-through cell for on-site measurement of pH, specific
040                conductance, temperature, dissolved oxygen and Eh (or pe)
040
040   Not Recommended: bailers
040
0418) Sampling Equipment Materials
041   ---------------------------
041
000 Info. in 042 applies when ground water is to be analysed for organics.
000 Info. in 043 applies when inorganics alone are to be analysed.
000
042   Recommended: Teflon
042
042   Acceptable: stainless steel
042
042   Not recommended: PVC, lucite, plexiglas, Vitron, silicone, neoprene,
042                    latex, natural or silicone rubber
042
043   Recommended: Teflon
043
043   Acceptable: stainless steel, PVC (unplasticized), polypropylene,
043               polyethylene
043
0449) Storage Containers
044   ------------------
044
045   - glass bottles for samples to be analysed for ammonia, sulfide and
045     ferrous iron
045
```

046 - polyethylene bottles for samples to be analysed for remaining inorganic
046 constituents
046
047 - glass bottles for samples to be analysed for organic chemicals; must
047 have Teflon-lined screw caps and Teflon septums
047
04810) Sampling Frequency
048 ----------
048
000 The info. given in 049 applies to detection programs only; info. in 050
000 applies to interpretive program
000
049 Recommended: consult the appropriate regulatory agency
049
049 Suggested: four times per year from all wells completed in deposits of
049 low permeability; monthly from all other wells
049
050 Suggested: once until complete monitoring network has been installed
050
05111) Analytical Program
051 ----------
051
051 Recommended: consult the appropriate regulatory agency
051
051 Suggested:
051
052 Initial sampling - "A" wells (plus "B" and "C" wells if feasible)
052
053 Initial sampling - "A" and "B" wells (plus "C" wells if feasible)
053
054 Sampling at "A" wells
054
000 List in 055 is extended group for municipal landfills
000
055 specific conductance, pH, temperature, Eh or pe, appearance, color,
055 turbidity
055 dissolved oxygen, chloride, iron, TSS, TDS, volatile solids,
055 sulfate, phosphate, alkalinity, acidity, nitrate-N, nitrite-N,
055 ammonia-N, calcium, magnesium, sodium, potassium, hardness, various
055 metals (Pb, Cu, Ni, Cr, Zn, Cd, Mn, Hg, As, Se, Ba, Ag), cyanide,
055 fluoride
055 COD, TOC, TOX, phenols, volatile acids, tannins and lignins, oil and
055 grease, organic-N
055 BOD, coliform bacteria (total, fecal, fecal streptococcus), standard
055 plate count
055
000 List in 056 is extended group for municipal sewage lagoons and ponds
000
056 specific conductance, pH, temperature, Eh or pe
056 dissolved oxygen, chloride, iron, nitrate, sodium, sulfate, phos-
056 phate, ammonium, manganese, boron, selenium, Zn, Cu, Pb
056 COD, TOC, TOX, organic-N, phenols
056 BOD
056

```
000 List in 057 is extended group for domestic septic systems and cesspools
000.
057        specific conductance, pH, temperature, Eh or pe
057        dissolved oxygen, chloride, nitrate, phosphate, ammonium, sulfate,
057        bicarbonate, iron
057        COD, TOC, organic-N
057        BOD
057
000 List in 058 is extended group for industrial wastewater lagoons and ponds
000
058        specific conductance, pH, temperature, Eh or pe, dissolved oxygen;
058        specific organic and/or inorganic constituents associated with the
058        lagoon or pond
058
000 List in 059 is extended group for industrial septic systems and cesspools
000
059        specific conductance, pH, temperature, Eh or pe, dissolved oxygen
059        chloride, nitrate, phosphate, ammonium, sulfate, bicarbonate, iron,
059        COD, TOC, organic-N
059        BOD
059        specific organic and/or inorganic chemicals produced or used at the
059        facility
059
000 List in 060 is extended group for industrial landfills where all wastes are
000 known
000
060        specific conductance, pH, temperature, Eh or pe
060        dissolved oxygen
060        specific organic and/or inorganic constituents associated with the
060        waste material
060
000 List in 061 is extended group for industrial landfills where all wastes
000 are not known
000
061        specific conductance, pH, temperature, Eh or pe, color, odor
061        dissolved oxygen, chloride, TDS, volatile solids, sulfate, acidity,
061        phosphate, alkalinity, nitrate-N, nitrite-N, ammonia-N, calcium,
061        magnesium, sodium, potassium, hardness, cyanide, fluoride, Pb, Cu,
061        Ni, Cr, Zn, Cd, Fe, Mn, Hg, As, Se, Ba, Ag
061        COD, TOC, TOX, phenols, volatile acids, tannins and lignins, oil and
061        and grease, organic-N, chloromethanes, methanol, acetone, benzene,
061        hexane
061
000 List in 062 is extended group for leaks and spills of crude oil and its
000 derivatives
000
062        specific conductance, pH, temperature, Eh or pe
062        dissolved oxygen
062        benzene, toluene, xylene (o-, m- and p-isomers)
062
000 List in 063 is extended group for leaks and spills of materials other than
000 crude oil and its derivatives
000
```

063 specific conductance, pH, temperature, Eh or pe
063 dissolved oxygen
063 specific organic and/or inorganic constituents associated with
063 the leaked or spilled material
063
000
064 Subsequent sampling - all "A", "B" and "C" wells
064
065 Sampling at "B" and "C" wells
065
000
000 Info. in 066 is key group for municipal landfills
000
066 specific conductance, pH, temperature, Eh or pe, dissolved oxygen,
066 chloride, iron, COD, TOC, TOX
066
000 Info. in 067 is key group for municipal sewage lagoons and ponds
000
067 specific conductance, pH, temperature, Eh or pe, dissolved oxygen,
067 chloride, iron, nitrate, sodium, sulfate, COD
067
000 Info. in 068 is key group for domestic septic systems and cesspools
000
068 specific conductance, pH, temperature, Eh or pe, dissolved oxygen,
068 chloride, nitrate, COD
068
000 Info. in 069 is key group for industrial wastewater lagoons and ponds
000
069 specific conductance, pH, temperature, Eh or pe, dissolved oxygen,
069 specific organic and/or inorganic constituents associated with
069 the lagoon or pond
069
000 Info. in 070 is key group for industrial septic systems and cesspools
000
070 specific conductance, pH, temperature, Eh or pe, dissolved oxygen,
070 chloride, nitrate, COD
070
000 Info. in 071 is key group for industrial landfills where all wastes are
000 known
000
071 specific conductance, pH, temperature, Eh or pe, dissolved oxygen,
071 specific organic and/or inorganic constituents associated with
071 the waste material
071
000 Info. in 072 is key group for industrial landfills where all wastes are
000 not known
000
072 specific conductance, pH, temperature, Eh or pe, dissolved oxygen,
072 chloride, COD, TOC, TOX
072
000 Info. in 073 is key group for leaks and spills of crude oil and its
000 derivatives
000

073 specific conductance, pH, temperature, Eh or pe, dissolved oxygen,
073 benzene, toluene, xylene (o-, m- and p-isomers)
073
000 Info. in 074 is key group for leaks and spills of materials other than
000 crude oil and its derivatives
000
074 - specific conductance, pH, temperature, Eh or pe, dissolved oxygen,
074 specific organic and/or inorganic constituents associated with the
074 leaked or spilled material
074
000
075 12) Sample Volumes and Sample Preservation
075 ----------
075
075 - consult with the manager of the lab that will be performing the
075 analytical work
075
000
076 ---------- End of Monitoring Strategy File ----------